

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
31 May 2001 (31.05.2001)

PCT

(10) International Publication Number
WO 01/38360 A2

(51) International Patent Classification⁷: **C07K 14/005**

Emeryville, CA 94608 (US). HOUGHTON, Michael [GB/US]; Chiron Corporation, 4560 Horton Street, R-440, Emeryville, CA 94608 (US).

(21) International Application Number: **PCT/US00/32326**

(74) Agents: HARBIN, Alisa, A.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 et al. (US).

(22) International Filing Date:
22 November 2000 (22.11.2000)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: English

(71) Applicant (*for all designated States except US*): CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(30) Priority Data:

60/167,502 24 November 1999 (24.11.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:

US 60/167,502 (CIP)
Filed on 24 November 1999 (24.11.1999)

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): COIT, Doris [US/US]; Chiron Corporation, 4560 Horton Street, R-440, Emeryville, CA 94608 (US). MEDINA-SELBY, Angelica [CL/US]; Chiron Corporation, 4560 Horton Street, R-440, Emeryville, CA 94608 (US). SELBY, Mark [US/US]; Chiron Corporation, 4560 Horton Street, R-440,

(73) Published:
Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/38360 A2

(54) Title: NOVEL HCV NON-STRUCTURAL POLYPEPTIDE

(57) Abstract: Polypeptides comprising a mutant non-structural Hepatitis C virus useful in diagnostic and/or immunogenic compositions are disclosed, in which the mutant is an N-terminal mutation that functionally disrupt the catalytic domain of NS3. Polynucleotides encoding these polypeptides, host cells transformed with polynucleotides and methods of using the polypeptides and polynucleotides are also disclosed.

5

NOVEL HCV NON-STRUCTURAL POLYPEPTIDE

FIELD OF THE INVENTION

The present invention relates to polypeptides comprising a mutant non-structural Hepatitis C virus ("HCV") polypeptide useful for immunogenic compounds for use against HCV, methods of preparing and using the same, and immunogenic compositions comprising the same. The present invention also relates to compositions comprising (a) a mutant non-structural HCV polypeptide and (b) a viral polypeptide that is not a non-structural HCV polypeptide and methods of using these compositions.

15

BACKGROUND OF THE INVENTION

HCV is now recognized as the major agent of chronic hepatitis and liver disease worldwide. It is estimated that HCV infects about 400 million people worldwide, corresponding to more than 3% of the world population.

20 Hepatitis C virus ("HCV") is a small enveloped RNA *flavivirus*, which contains a positive-stranded RNA genome of about 10 kilobases. The genome has a single uninterrupted ORF that encodes a protein of 3010-3011 amino acids. The structural proteins of HCV include a core protein (C), which is highly immunogenic, as well as two envelope proteins (E1 and E2), which likely form a heterodimer *in vivo*, and non-structural proteins NS2-NS5. It is known that the NS3 region of the virus is important 25 for post-translational processing of the polyprotein into individual proteins, and the NS5 region encodes an RNA-dependant RNA polymerase.

30 Virus-specific T lymphocytes, along with neutralizing antibodies, are the mainstay of the antiviral immune defense in established viral infections. Whereas CD8⁺ cytotoxic T cells eliminate virus-infected-cells, CD4⁺ T helper cells are essential for the efficient regulation of the antiviral immune response. CD4⁺ T helper cells recognize specific antigens as peptides bound to autologous HLA class II molecules (viral antigens or particles are taken up by professional antigen-presenting cells, processed to peptides, bound to HLA class II molecules in the lysosomal compartment,

and transported back to the cell surface). Several observations support an important role of CD4⁺ T cells in the elimination of HCV infection. Tsai *et al.*, 1997 Hepatology 25:449-458; Diepolder et al 1995 Lancet 346: 1—6-1009; Missale et al 1996 JCI 98: 706-714; Botarelli et al 1993; Gastro 104: 580-587; Diepolder et al 1997 J. Virol 71: 6011. Immunogenic peptides usually have a minimal length of 8-11 amino acids. However, since the peptide binding groove of HLA class II molecules seems to be open at both ends, longer peptides are tolerated. Thus peptides eluted from HLA class II molecules are typically in the range of 15-25 amino acids. HLA class II molecules are extremely polymorphic and each allele seems to have its individual requirements for peptide binding. Thus the HLA class II repertoire of a given individual determines which viral peptides can be presented to T cells. Recognition of the specific HLA-peptide complex by the T cell receptor accompanied by appropriate costimulatory signals lead to T cell activation, secretion of cytokines, and T cell proliferation.

Numerous studies demonstrate that HLA Class II restricted CD4⁺ responses are determined by stimulating peripheral blood mononuclear cells with recombinant viral antigens or peptides. Botarelli *et al.*, (1993) Gastroenterology 104:580-587; Farrari *et al.*, (1994) Hepatology 19:286-295; Minutello *et al.*, (1993) C. J. Exp. Med. 178:17-25; Hoffmann *et al.*, (1995) Hepatology 21:632-638; Iwata *et al.*, (1995) Hepatology 22:1057-1064; and Tsai.*et al.*, (1995) Hepatology 21:908-912.

Polyclonal multispecific CD8⁺ T cell responses have been detected in patients with chronic hepatitis C. Additionally, CD8⁺ CTL's were shown to be important in resolving acute HCV infection in chimpanzees (Cooper *et al.*, Immunity 1999). About 50% of patients with chronic hepatitis C demonstrate a detectable virus-specific CD4⁺ T cell response, which is most frequently directed against HCV core and/or NS4 and tends to be more common in patients who achieve sustained viral clearance during interferon- α therapy.

Depending on the pattern of lymphokines, CD4⁺ T helper cells have been classified as TH1, TH0, or TH2. Cytokines of the TH1 type are typically IFN- γ , lymphotoxin, and interleukin-2 (IL-2), which are believed to support activation of virus-specific CD8⁺ T cells and natural killer cells. The TH2 cytokines IL-4, IL-5, IL-10, and IL-13 are important for B cell activation and differentiation, thus inducing a humoral immune response.

During acute hepatitis C infection a strong and sustained TH1/TH0 response to NS3 and possibly to other nonstructural proteins is associated with a self-limited course of the disease. Diapolder *et al.*, (1995) Lancet 346:1006-1007, showed all CD4⁺ T cell clones to have a TH1 or TH0 cytokine profile, suggesting that the clones support 5 cytotoxic immune mechanisms *in vivo*. The majority of CD4⁺ T cell clones responded to a relatively short segment of NS3, namely amino acids 1207-1278, suggesting that this region of NS3 is immunodominant for CD4⁺ T cells. More than 70% of those who contract HCV develop chronic infection and hepatitis, and a significant portion of them progress to cirrhosis and eventually hepatocellular carcinoma. The only approved 10 therapy at present is a 6- to 12- month course of interferon α , which leads to sustained improvement in only 20% of patients. So far, no commercial vaccine is available.

Thus, there remains a need for compositions and methods capable of promoting anti-HCV responses.

15 SUMMARY OF THE INVENTION

In one aspect, the present invention relates to isolated polypeptides comprising mutant hepatitis C ("HCV") polypeptides comprising at least portions of NS3, NS4, and NS5. In a preferred aspect, NS3 is encoded by a nucleic acid sequence having an N-terminal deletion to remove the catalytic domain. The NS mutant polypeptides can 20 include NS3, NS4s, NS4b, NS5a, NS5b or portions thereof. For example, in various embodiments, the mutant NS polypeptide comprises NS3, NS4 (NS4a and NS4b) and NS5 (NS5a and NS5b). In other embodiments, the NS polypeptide consists of NS3 and NS4 (for example, NS4a and/or NS4b) or NS3 and NS5 (for example, NS5a and/or NS5b). Other combinations of full-length or fragments of non-structural components 25 are also contemplated.

In another preferred aspect, the polypeptides further comprise a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, 30 E1 or E2 of HCV. Such polypeptides need not be encoded by a natural HCV genome, and include, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV.

Thus, the invention includes an isolated mutant non-structural ("NS") HCV polypeptide comprising a polypeptide having a mutation in the catalytic domain of NS3 that functionally disrupts the catalytic domain. The mutation can be, for example, a deletion or a substitution mutation. In certain embodiments, the mutant NS polypeptide 5 comprises NS3, NS4 and NS5. In other embodiments, the mutant NS polypeptides described herein further comprise a second viral polypeptide that is not NS3, NS4, or NS5 of HCV, for example an HCV Core polypeptide ("C"), or fragment thereof, or an HCV envelope protein ("E"), for example E1 and/or E2. In certain embodiments, C is truncated (*e.g.*, at amino acid 121).

10 In another aspect, the present invention relates to compositions comprising any of the mutant hepatitis C ("HCV") polypeptides described herein, for example polypeptides comprising at least portions of NS3, NS4, and NS5. In a preferred aspect, NS3 is encoded by a nucleic acid sequence having an N-terminal deletion to disrupt the function of the catalytic domain, for example by removing this domain. In another 15 preferred aspect, the polypeptides further comprise a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by a natural HCV genome, and include, for example, 20 truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV. In another aspect, the invention includes a composition comprising (a) any of the polypeptides described herein; and (b) a pharmaceutically acceptable excipient (*e.g.*, carrier and/or adjuvant).

25 In another aspect, the invention includes an isolated and purified polynucleotide which encodes any of the mutant HCV polypeptides described herein. In certain embodiments, the invention includes a composition comprising (a) the isolated purified polynucleotide encoding any of the mutant HCV polypeptides; and (b) a pharmaceutically acceptable excipient. The polynucleotide, can be for example, DNA in a plasmid, or is in a plasmid. Additionally, the polynucleotides described herein may 30 be included in an expression vector as shown in the attached Figures and Sequence Listings.

In another aspect, the present invention relates to host cells transformed with expression vectors comprising a nucleic acid sequence encoding a mutant HCV polypeptide comprising at least portions of NS3, NS4, and NS5. In a preferred aspect, the expression vectors of the host cells further comprises at least one nucleic acid sequence encoding a viral polypeptide that is not a non-structural HCV polypeptide.

Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by a natural HCV genome, and include, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV. In another preferred aspect the nucleic acid sequences of the expression vectors are coexpressed. In yet another preferred aspect, the host cells are yeast cells or mammalian cells.

In another aspect, the present invention relates to expression vectors comprising a nucleic acid sequence encoding a mutant HCV polypeptide comprising NS3, NS4, and NS5. In a preferred aspect, the expression vectors of the host cells further comprises at least one nucleic acid sequence encoding a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV.

Importantly, such polypeptides need not be encoded by a natural HCV genome, such as, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV. In another aspect, the present invention relates to methods of preparing a mutant HCV polypeptides. In a preferred aspect, the method comprises the steps of transforming a host cell with an expression vector, said vector comprising a nucleic acid sequence encoding a mutant HCV polypeptide comprising at least portions of NS3, NS4, and NS5, and isolating said polypeptide. In another preferred aspect the HCV polypeptide further comprises a viral polypeptide that is not a non-structural HCV polypeptide.

Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by

a natural HCV genome, and include, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV. In another preferred aspect the host cells are yeast cells or mammalian cells.

- 5 In another aspect, the present invention relates to antibodies which specifically bind to mutant HCV polypeptide comprising NS3, NS4, and NS5, and to methods of making and using the same. In a preferred aspect, the HCV polypeptide further comprises a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, 10 truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by a natural HCV genome, such as, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, and include, for example, polypeptides of HBV. In another preferred aspect, the antibody is either monoclonal or 15 polyclonal.

In yet another aspect, a method of preparing a mutant NS HCV polypeptide, wherein the method comprises the steps of (a) transforming a host cell with any of the expression vectors described herein, under conditions wherein the polypeptide is expressed; and (b) isolating the polypeptide. The host cell can be, for example, a yeast 20 cell, a mammalian cell a plant cell or an insect cell. The polypeptide can be expressed and isolated intracellularly or can be secreted and isolated from the surrounding environment.

In a still further aspect, a method of eliciting an immune response in a subject is provided. The immune response can be elicited by administering any of the 25 polynucleotides and/or polypeptides described herein in one or multiple doses.

These and other embodiments of the subject invention will readily occur to those of skill in the art in light of the disclosure herein.

BRIEF DESCRIPTION OF THE FIGURES

- 30 FIG. 1 shows the cloning scheme for generating pCMV-NS35.
FIG. 2 shows the 9621bp vector pCMV-NS35.

FIG. 3 shows the nucleic acid sequence of pCMV-NS35 (SEQ ID NO:1), including the nucleic acid sequence of the NS35 ORF, and also the translation of NS35 (SEQ ID NO:2).

FIG. 4 shows the 9621bp pCMV-delNS35.

5 FIG. 5 shows the nucleic acid sequence of pCMV-delNS35 (SEQ ID NO:3), including the nucleic acid sequence of the delNS35 ORF, and also the translation of the delNS35 polypeptide (SEQ ID NO:4).

FIG. 6 shows the 4276bp pCMV-II.

FIG. 7 shows the nucleic acid sequence of pCMV-II (SEQ ID NO:5).

10 FIG. 8 shows the 6300bp pCMV-NS34A.

FIG. 9 shows the nucleic acid sequence of pCMV-NS34A (SEQ ID NO:6), including the nucleic acid sequence of the NS34A ORF, and also the translation of NS34A (SEQ ID NO:7).

FIG. 10 shows the cloning scheme for generating pd.ΔNS3NS5.

15 FIG. 11 shows the nucleic and amino acid sequences of pd.ΔNS3NS5 (SEQ ID NO:8 and 9).

FIG. 12 shows the Western blot of proteins expressed by *S. cerevisiae* strain AD3 transformed with pd.ΔNS3NS5.

FIG. 13 shows the cloning scheme for generating pd.ΔNS3NS5.pj.

20 FIG. 14 shows the nucleic and amino acid sequences of pd.ΔNS3NS5.pj (SEQ ID NO:10 and 11).

FIG. 15 shows the Western blot of proteins expressed by *S. cerevisiae* strain AD3 transformed with pd.ΔNS3NS5.pj, specifically demonstrating the expression of ΔNS3NS5 polypeptide.

25 FIG. 16 shows the cloning scheme for generating pd.ΔNS3NS5.pj.core121RT and pd.ΔNS3NS5.pj.core173RT.

FIG. 17 shows the nucleic and amino acid sequences of pd.ΔNS3NS5.pj.core121 (SEQ ID NO:12 and 13).

FIG. 18 shows the nucleic and amino acid sequences of pd.ΔNS3NS5.pj.core173 (SEQ ID NO:14 and 15).

30 FIG. 19 shows the Western blot of proteins expressed by *S. cerevisiae* strain AD3 transformed with pd.ΔNS3NS5.pj, specifically demonstrating the expression of

Δ NS3NS5.core121 and Δ NS3NS5.core173 polypeptides. Lanes 1 and 7 show See Blue Standards. Lane 2 shows control yeast plasmid. Lanes 3 and 4 show Δ NS3NS5.core121RT polypeptide, colonies 1 and 2. Lanes 5 and 6 show Δ NS3NS5.core173RT polypeptide, colonies 3 and 4.

5 FIG. 20 shows the cloning scheme for generating pd Δ NS3NS5.pj.core140RT and pd Δ NS3NS5.pj.core150RT.

FIG. 21 shows the nucleic and amino acid sequences of pd. Δ NS3NS5.pj.core140 (SEQ ID NO:16 and 17).

10 FIG. 22 shows the nucleic and amino acid sequences of pd. Δ NS3NS5.pj.core150 (SEQ ID NO:18 and 19).

FIG. 23 shows the Western blot of proteins expressed by *S. cerevisiae* strain AD3 transformed with pd. Δ NS3NS5.pj, specifically demonstrating the expression of Δ NS3NS5core140 and Δ NS3NS5core150 polypeptides. Lane 1 shows See Blue Standards. Lanes 2 and 3 show Δ NS3NS5core140RT polypeptide, colonies 5 and 6.

15 Lanes 4 and 5 show Δ NS3NS5core150RT polypeptide, colonies 7 and 8. Lane 6 shows control yeast plasmid. Lane 7 shows Δ NS3NS5core121RT polypeptide, colony 1. Lane 8 shows Δ NS3NS5core173RT polypeptide, colony 5.

DETAILED DESCRIPTION OF THE INVENTION

20 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA techniques, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL (1989); DNA CLONING, VOLUMES I AND II (D. N. Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M. J. Gait ed., 1984); NUCLEIC ACID HYBRIDIZATION (B. D. Hames & S. J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B. D. Hames & S. J. Higgins eds. 1984); ANIMAL CELL CULTURE (R. I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS OF ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J. H. Miller and M. P. Calos eds. 1987, Cold Springs Harbor Laboratory), Methods in Enzymology Vol.

154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively); Mayer and Walker
eds. (1987), IMMUNOHISTOCHEMICAL METHODS IN CELL AND
MOLECULAR BIOLOGY (Academic Press, London); Scopes, (1987), PROTEIN
PURIFICATION: PRINCIPALS AND PRACTICE, Second Edition (Springer-Verlag,
5 New York); and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES
I-IV (D. M. Weir and C. C. Blackwell eds. 1986).

It must be noted that, as used in this specification and the appended claims, the
singular forms "a", "an" and "the" include plural referents unless the content clearly
dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of
10 two or more antigens, and the like.

I. Definitions

In describing the present invention, the following terms will be employed, and
are intended to be defined as indicated below.

15 The term "hepatitis C virus" (HCV) refers to an agent causative of Non-A, Non-
B Hepatitis (NANBH). The nucleic acid sequence and putative amino acid sequence of
HCV is described in U.S. Patent Nos. 5,856,437 and 5,350,671. The disease caused by
HCV is called hepatitis C, formerly called NANBH. The term HCV, as used herein,
denotes a viral species of which pathogenic strains cause NANBH, as well as
20 attenuated strains or defective interfering particles derived therefrom.

HCV is a member of the viral family flaviviridae. The morphology and
composition of Flavivirus particles are known, and are discussed in Reed et al., *Curr.*
Stud. Hematol. Blood Transfus. (1998), 62:1-37; HEPATITIS C VIRUSES IN FIELDS
VIROLOGY (B.N. Fields, D.M. Knipe, P.M. Howley, eds.) (3d ed. 1996). It has
25 recently been found that portions of the HCV genome are also homologous to
pestiviruses. Generally, with respect to morphology, Flaviviruses contain a central
nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter
of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer
surface of the virion envelope are projections that are about 5-10 nm long with terminal
30 knobs about 2 nm in diameter.

The HCV genome is comprised of RNA. It is known that RNA containing
viruses have relatively high rates of spontaneous mutation. Therefore, there can be

multiple strains, which can be virulent or avirulent, within the HCV class or species.

The ORF of HCV, including the translation spans of the core, non-structural, and envelope proteins, is shown in U.S. Patent Nos. 5,856,437 and 5,350,671.

The terms "polypeptide" and "protein" refer to a polymer of amino acid residues and are not limited to a minimum length of the product. Thus, peptides, oligopeptides, dimers, multimers, and the like, are included within the definition. Both full-length proteins and fragments thereof are encompassed by the definition. The terms also include postexpression modifications of the polypeptide, for example, glycosylation, acetylation, phosphorylation and the like. Furthermore, for purposes of the present invention, a "polypeptide" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the desired activity. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the proteins or errors due 10 to PCR amplification.

An HCV polypeptide is a polypeptide, as defined above, derived from the HCV polyprotein. The polypeptide need not be physically derived from HCV, but may be synthetically or recombinantly produced. Moreover, the polypeptide may be derived from any of the various HCV strains, such as from strains 1, 2, 3 or 4 of HCV. A 20 number of conserved and variable regions are known between these strains and, in general, the amino acid sequences of epitopes derived from these regions will have a high degree of sequence homology, e.g., amino acid sequence homology of more than 30%, preferably more than 40%, when the two sequences are aligned and homology determined by any of the programs or algorithms described herein. Thus, for example, 25 the term "NS4" polypeptide refers to native NS4 from any of the various HCV strains, as well as NS4 analogs, muteins and immunogenic fragments, as defined further below.

Further, the terms " Δ NS35," "delNS35," " Δ NS3NS5," and " Δ NS3-5" as used herein refer to a mutant polypeptide, comprising at least portions of NS3, NS4, or NS5, comprising a deletion in, or mutation of, the NS3 protease active site region to render 30 the protease non-functional. In one embodiment, Δ NS3-5 comprises amino acids 1242-3011, as shown in FIG. 5, or polypeptides substantially homologous thereto. It will be readily apparent to one of ordinary skill in the art how to determine that NS3 protease

has been rendered non-functional. If the protease is functional, one will obtain protein of the expected molecular weight upon expression. As set forth in Example 2 and Figure 15, using SDS-page, 4-20%, a protein having a molecular weight of approximately 194kD was obtained when strain AD3 was transformed with
5 pd. Δ NS3NS5.PJ clone #5. One skilled in the art could readily determine whether a protein of the desired molecular weight was expressed for any given deletion or mutation.

The terms "analog" and "mutein" refer to biologically active derivatives of the reference molecule, or fragments of such derivatives, that retain desired activity, such
10 as the ability to stimulate a cell-mediated immune response, as defined below. In general, the term "analog" refers to compounds having a native polypeptide sequence and structure with one or more amino acid additions, substitutions (generally conservative in nature) and/or deletions, relative to the native molecule, so long as the modifications do not destroy immunogenic activity. The term "mutein" refers to
15 peptides having one or more peptide mimics ("peptoids"), such as those described in International Publication No. WO 91/04282. Preferably, the analog or mutein has at least the same immunoactivity as the native molecule. Methods for making polypeptide analogs and muteins are known in the art and are described further below.

Particularly preferred analogs include substitutions that are conservative in
20 nature, i.e., those substitutions that take place within a family of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cysteine, serine
25 threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids. For example, it is reasonably predictable that an isolated replacement of leucine with isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid, will not have a major effect on the biological activity.
30 For example, the polypeptide of interest may include up to about 5-10 conservative or non-conservative amino acid substitutions, or even up to about 15-25 conservative or non-conservative amino acid substitutions, or any integer between 5-25, so long as the

desired function of the molecule remains intact. One of skill in the art may readily determine regions of the molecule of interest that can tolerate change by reference to Hopp/Woods and Kyte-Doolittle plots, well known in the art.

By "fragment" is intended a polypeptide consisting of only a part of the intact full-length polypeptide sequence and structure. The fragment can include a C-terminal deletion and/or an N-terminal deletion of the native polypeptide. An "immunogenic fragment" of a particular HCV protein will generally include at least about 5-10 contiguous amino acid residues of the full-length molecule, preferably at least about 15-25 contiguous amino acid residues of the full-length molecule, and most preferably 10 at least about 20-50 or more contiguous amino acid residues of the full-length molecule, that define an epitope, or any integer between 5 amino acids and the full-length sequence, provided that the fragment in question retains immunogenic activity, as measured by the assays described herein. For a description of various HCV epitopes, see, e.g., Chien et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:10011-10015; 15 Chien et al., *J. Gastroent. Hepatol.* (1993) 8:S33-39; Chien et al., International Publication No. WO 93/00365; Chien, D.Y., International Publication No. WO 94/01778; commonly owned, allowed U.S. Patent Application Serial Nos. 08/403,590 and 08/444,818.

The term "epitope" as used herein refers to a sequence of at least about 3 to 5, 20 preferably about 5 to 10 or 15, and not more than about 1,000 amino acids (or any integer therebetween), which define a sequence that by itself or as part of a larger sequence, binds to an antibody generated in response to such sequence. There is no critical upper limit to the length of the fragment, which may comprise nearly the full-length of the protein sequence, or even a fusion protein comprising two or more 25 epitopes from the HCV polyprotein. An epitope for use in the subject invention is not limited to a polypeptide having the exact sequence of the portion of the parent protein from which it is derived. Indeed, viral genomes are in a state of constant flux and contain several variable domains which exhibit relatively high degrees of variability between isolates. Thus the term "epitope" encompasses sequences identical to the 30 native sequence, as well as modifications to the native sequence, such as deletions, additions and substitutions (generally conservative in nature).

Regions of a given polypeptide that include an epitope can be identified using any number of epitope mapping techniques, well known in the art. See, e.g., *Epitope Mapping Protocols* in Methods in Molecular Biology, Vol. 66 (Glenn E. Morris, Ed., 1996) Humana Press, Totowa, New Jersey. For example, linear epitopes may be determined by e.g., concurrently synthesizing large numbers of peptides on solid supports, the peptides corresponding to portions of the protein molecule, and reacting the peptides with antibodies while the peptides are still attached to the supports. Such techniques are known in the art and described in, e.g., U.S. Patent No. 4,708,871; Geysen et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:3998-4002; Geysen et al. (1986) *Molec. Immunol.* 23:709-715. Similarly, conformational epitopes are readily identified by determining spatial conformation of amino acids such as by, e.g., x-ray crystallography and 2-dimensional nuclear magnetic resonance. See, e.g., *Epitope Mapping Protocols*, *supra*. Antigenic regions of proteins can also be identified using standard antigenicity and hydropathy plots, such as those calculated using, e.g., the Omiga version 1.0 software program available from the Oxford Molecular Group. This computer program employs the Hopp/Woods method, Hopp et al., *Proc. Natl. Acad. Sci USA* (1981) 78:3824-3828 for determining antigenicity profiles, and the Kyte-Doolittle technique, Kyte et al., *J. Mol. Biol.* (1982) 157:105-132 for hydropathy plots.

As used herein, the term "conformational epitope" refers to a portion of a full-length protein, or an analog or mutein thereof, having structural features native to the amino acid sequence encoding the epitope within the full-length natural protein. Native structural features include, but are not limited to, glycosylation and three dimensional structure. Preferably, a conformational epitope is produced recombinantly and is expressed in a cell from which it is extractable under conditions which preserve its desired structural features, e.g. without denaturation of the epitope. Such cells include bacteria, yeast, insect, and mammalian cells. Expression and isolation of recombinant conformational epitopes from the HCV polyprotein are described in e.g., International Publication Nos. WO 96/04301, WO 94/01778, WO 95/33053, WO 92/08734.

An "immunological response" to an HCV antigen (including both polypeptide and polynucleotides encoding polypeptides that are expressed *in vivo*) or composition is the development in a subject of a humoral and/or a cellular immune response to molecules present in the composition of interest. For purposes of the present invention,

a "humoral immune response" refers to an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTLs"). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the intracellular destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376; and the examples below.

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or the production or activation of helper T- cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the activation of suppressor T-cells and/or $\gamma\delta$ T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate

antibody-complement, or antibody dependent cell cytotoxicity (ADCC) to provide protection or alleviation of symptoms to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

5 A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at
10 the 3' (carboxy) terminus. A transcription termination sequence may be located 3' to the coding sequence.

A "nucleic acid" molecule or "polynucleotide" can include both double- and single-stranded sequences and refers to, but is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral (e.g. DNA
15 viruses and retroviruses) or procaryotic DNA, and especially synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their desired function. Thus,
20 a given promoter operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper transcription factors, etc., are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding
25 sequence, as can transcribed introns, and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, viral, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation is not associated with all or a portion of the
30 polynucleotide with which it is associated in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. In general, the gene of interest is cloned and then

expressed in transformed organisms, as described further below. The host organism expresses the foreign gene to produce the protein under expression conditions.

A "control element" refers to a polynucleotide sequence which aids in the expression of a coding sequence to which it is linked. The term includes promoters, 5 transcription termination sequences, upstream regulatory domains, polyadenylation signals, untranslated regions, including 5'-UTRs and 3'-UTRs and when appropriate, leader sequences and enhancers, which collectively provide for the transcription and translation of a coding sequence in a host cell.

A "promoter" as used herein is a DNA regulatory region capable of binding 10 RNA polymerase in a host cell and initiating transcription of a downstream (3' direction) coding sequence operably linked thereto. For purposes of the present invention, a promoter sequence includes the minimum number of bases or elements necessary to initiate transcription of a gene of interest at levels detectable above background. Within the promoter sequence is a transcription initiation site, as well as 15 protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eucaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

A control sequence "directs the transcription" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding 20 sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

"Expression cassette" or "expression construct" refers to an assembly which is capable of directing the expression of the sequence(s) or gene(s) of interest. The expression cassette includes control elements, as described above, such as a promoter 25 which is operably linked to (so as to direct transcription of) the sequence(s) or gene(s) of interest, and often includes a polyadenylation sequence as well. Within certain embodiments of the invention, the expression cassette described herein may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include, one or more selectable markers, a signal which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 30 origin of replication), at least one multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

"Transformation," as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for insertion: for example, transformation by direct uptake, transfection, infection, and the like. For particular methods of transfection, see further below. The exogenous polynucleotide 5 may be maintained as a nonintegrated vector, for example, an episome, or alternatively, may be integrated into the host genome.

A "host cell" is a cell which has been transformed, or is capable of transformation, by an exogenous DNA sequence.

By "isolated" is meant, when referring to a polypeptide, that the indicated 10 molecule is separate and discrete from the whole organism with which the molecule is found in nature or is present in the substantial absence of other biological macromolecules of the same type. The term "isolated" with respect to a polynucleotide is a nucleic acid molecule devoid, in whole or part, of sequences normally associated with it in nature; or a sequence, as it exists in nature, but having heterologous sequences in 15 association therewith; or a molecule disassociated from the chromosome.

The term "purified" as used herein preferably means at least 75% by weight, more preferably at least 85% by weight, more preferably still at least 95% by weight, and most preferably at least 98% by weight, of biological macromolecules of the same type are present.

20 "Homology" refers to the percent identity between two polynucleotide or two polypeptide moieties. Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 50%, preferably at least about 75%, more preferably at least about 80%-85%, preferably at least about 90%, and most preferably at least about 95%-98%, or more, sequence identity over a 25 defined length of the molecules. As used herein, substantially homologous also refers to sequences showing complete identity to the specified DNA or polypeptide sequence. The term "substantially homologous" as used herein in reference to Δ NS35 generally refers to an HCV nucleic or amino acid sequence that is at least 60% identical to the entire sequence of the polypeptide encoded by Δ NS35 (see FIG. 5), where the sequence 30 identity is preferably at least 75%, more preferably at least 80%, still more preferably at least about 85%, especially more than about 90%, most preferably 95% or greater, particularly 98% or greater. These homologous polypeptides include fragments,

including mutants and allelic variants of the fragments. Identity between the two sequences is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1. Thus, for 5 example, the present invention includes an isolate which is 80% identical to a polypeptide encoded by ΔNS35. In some aspects of the invention, the polypeptide of the present invention is substantially homologous to the ΔNS35.

In general, "identity" refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, 10 respectively. Percent identity can be determined by a direct comparison of the sequence information between two molecules by aligning the sequences, counting the exact number of matches between the two aligned sequences, dividing by the length of the shorter sequence, and multiplying the result by 100. Readily available computer programs can be used to aid in the analysis, such as ALIGN, Dayhoff, M.O. in *Atlas of 15 Protein Sequence and Structure* M.O. Dayhoff ed., 5 Suppl. 3:353-358, National biomedical Research Foundation, Washington, DC, which adapts the local homology algorithm of Smith and Waterman *Advances in Appl. Math.* 2:482-489, 1981 for peptide analysis. Programs for determining nucleotide sequence identity are available in the Wisconsin Sequence Analysis Package, Version 8 (available from Genetics 20 Computer Group, Madison, WI) for example, the BESTFIT, FASTA and GAP programs, which also rely on the Smith and Waterman algorithm. These programs are readily utilized with the default parameters recommended by the manufacturer and described in the Wisconsin Sequence Analysis Package referred to above. For example, percent identity of a particular nucleotide sequence to a reference sequence 25 can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions.

Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by 30 IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a

gap of six). From the data generated the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be
5 used using the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at the following internet address:
10 <http://www.ncbi.nlm.gov/cgi-bin/BLAST>.

Alternatively, homology can be determined by hybridization of polynucleotides under conditions which form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. DNA sequences that are substantially homologous can be
15 identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., *supra*; *DNA Cloning, supra*; *Nucleic Acid Hybridization, supra*.

"Stringency" refers to conditions in a hybridization reaction that favor
20 association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated Tm of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized
25 to the sequence of interest and then washed under conditions of different stringencies. See Sambrook et al. at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a
30 magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10⁻⁹ to 10⁻⁸ g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a

smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a
5 conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate
10 conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:
15 $T_m = 81 + 16.6(\log_{10}C_i) + 0.4[(G + C)] - 0.6(\% \text{formamide}) - 600/n - 1.5(\%\text{mismatch})$. where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284). In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with 95% to 100% homologous to the target
20 fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background
25 are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

By "nucleic acid immunization" is meant the introduction of a nucleic acid molecule encoding one or more selected antigens into a host cell, for the *in vivo*
30 expression of the antigen or antigens. The nucleic acid molecule can be introduced directly into the recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have

been removed from the host. In the latter case, the transformed cells are reintroduced into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

- An "open reading frame" or ORF is a region of a polynucleotide sequence
5 which encodes a polypeptide; this region can represent a portion of a coding sequence or a total coding sequence.

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which comprise at least one antigen binding site. An "antigen binding site" is formed from the folding of the variable domains of an antibody molecule(s) to
10 form three-dimensional binding sites with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows specific binding to form an antibody-antigen complex. An antigen binding site may be formed from a heavy- and/or light-chain domain (VH and VL, respectively), which form hypervariable loops which contribute to antigen binding. The term "antibody"
15 includes, without limitation, polyclonal antibodies, monoclonal antibodies, chimeric antibodies, altered antibodies, univalent antibodies, Fab proteins, and single-domain antibodies. In many cases, the binding phenomena of antibodies to antigens is equivalent to other ligand/anti-ligand binding.

If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit,
20 goat, horse, etc.) is immunized with an immunogenic polypeptide bearing an HCV epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an HCV epitope contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal
25 antisera are known in the art, see for example, Mayer and Walker, eds. (1987)
IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY
(Academic Press, London).

Monoclonal antibodies directed against HCV epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal
30 antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g.,

M. Schreier et al. (1980) HYBRIDOMA TECHNIQUES; Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS; Kennett et al. (1980) MONOCLONAL ANTIBODIES; see also, U.S. Pat. Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of 5 monoclonal antibodies produced against HCV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc. As used herein, a "single domain antibody" (dAb) is an antibody which is comprised of an HL domain, which binds specifically with a designated antigen. A dAb does not contain a VL domain, but may contain other antigen binding domains known to exist to antibodies, for example, the 10 kappa and lambda domains. Methods for preparing dabs are known in the art. See, for example, Ward et al, Nature 341: 544 (1989).

Antibodies can also be comprised of VH and VL domains, as well as other known antigen binding domains. Examples of these types of antibodies and methods for their preparation and known in the art (see, e.g., U.S. Pat. No. 4,816,467), and 15 include the following. For example, "vertebrate antibodies" refers to antibodies which are tetramers or aggregates thereof, comprising light and heavy chains which are usually aggregated in a "Y" configuration and which may or may not have covalent linkages between the chains. In vertebrate antibodies, the amino acid sequences of the chains are homologous with those sequences found in antibodies produced in 20 vertebrates, whether *in situ* or *in vitro* (for example, in hybridomas). Vertebrate antibodies include, for example, purified polyclonal antibodies and monoclonal antibodies, methods for the preparation of which are described infra.

"Hybrid antibodies" are antibodies where chains are separately homologous with reference to mammalian antibody chains and represent novel assemblies of them, 25 so that two different antigens are precipitable by the tetramer or aggregate. In hybrid antibodies, one pair of heavy and light chains are homologous to those found in an antibody raised against a first antigen, while a second pair of chains are homologous to those found in an antibody raised against a second antibody. This results in the property of "divalence", i.e., the ability to bind two antigens simultaneously. Such hybrids can 30 also be formed using chimeric chains, as set forth below.

"Chimeric antibodies" refers to antibodies in which the heavy and/or light chains are fusion proteins. Typically, one portion of the amino acid sequences of the

chain is homologous to corresponding sequences in an antibody derived from a particular species or a particular class, while the remaining segment of the chain is homologous to the sequences derived from another species and/or class. Usually, the variable region of both light and heavy chains mimics the variable regions of antibodies derived from one species of vertebrates, while the constant portions are homologous to the sequences in the antibodies derived from another species of vertebrates. However, the definition is not limited to this particular example. Also included is any antibody in which either or both of the heavy or light chains are composed of combinations of sequences mimicking the sequences in antibodies of different sources, whether these 5 sources be from differing classes or different species of origin, and whether or not the fusion point is at the variable/constant boundary. Thus, it is possible to produce 10 antibodies in which neither the constant nor the variable region mimic known antibody sequences. It then becomes possible, for example, to construct antibodies whose 15 variable region has a higher specific affinity for a particular antigen, or whose constant region can elicit enhanced complement fixation, or to make other improvements in properties possessed by a particular constant region.

Another example is "altered antibodies", which refers to antibodies in which the naturally occurring amino acid sequence in a vertebrate antibody has been varied. Utilizing recombinant DNA techniques, antibodies can be redesigned to obtain desired 20 characteristics. The possible variations are many, and range from the changing of one or more amino acids to the complete redesign of a region, for example, the constant region. Changes in the constant region, in general, to attain desired cellular process characteristics, e.g., changes in complement fixation, interaction with membranes, and other effector functions. Changes in the variable region can be made to alter antigen 25 binding characteristics. The antibody can also be engineered to aid the specific delivery of a molecule or substance to a specific cell or tissue site. The desired alterations can be made by known techniques in molecular biology, e.g., recombinant techniques, site-directed mutagenesis, etc.

Yet another example are "univalent antibodies", which are aggregates 30 comprised of a heavy-chain/light-chain dimer bound to the Fc (i.e., stem) region of a second heavy chain. This type of antibody escapes antigenic modulation. See, e.g., Glennie et al. Nature 295: 712 (1982). Included also within the definition of antibodies

- are "Fab" fragments of antibodies. The "Fab" region refers to those portions of the heavy and light chains which are roughly equivalent, or analogous, to the sequences which comprise the branch portion of the heavy and light chains, and which have been shown to exhibit immunological binding to a specified antigen, but which lack the effector Fc portion. "Fab" includes aggregates of one heavy and one light chain (commonly known as Fab'), as well as tetramers containing the 2H and 2L chains (referred to as F(ab)2), which are capable of selectively reacting with a designated antigen or antigen family. Fab antibodies can be divided into subsets analogous to those described above, i.e., "vertebrate Fab", "hybrid Fab", "chimeric Fab", and "altered Fab".
- 5 Methods of producing Fab fragments of antibodies are known within the art and include, for example, proteolysis, and synthesis by recombinant techniques.

"Antigen-antibody complex" refers to the complex formed by an antibody that is specifically bound to an epitope on an antigen.

- 10 "Immunogenic polypeptide" refers to a polypeptide that elicits a cellular and/or humoral immune response in a mammal, whether alone or linked to a carrier, in the presence or absence of an adjuvant.

"Antigenic determinant" refers to the site on an antigen or hapten to which a specific antibody molecule or specific cell surface receptor binds.

- 15 As used herein, "treatment" refers to any of (i) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may be effected prophylactically (prior to infection) or therapeutically (following infection).

- 20 By "vertebrate subject" is meant any member of the subphylum cordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The

invention described herein is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

II. Modes of Carrying out the Invention

- 5 Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.
- 10 Although a number of compositions and methods similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

General Overview

- 15 An aim of an HCV vaccine is to generate broad immunity to a wide breadth of antigens because HCV is so divergent and because humoral as well as cellular immune responses are desirable to combat this human pathogen. While antibodies generated against the envelope glycoprotein(s) might aid in virus neutralization, there is additional benefit to be derived from a vaccine that includes other regions. The
- 20 likelihood of T-helper responses generated against a polypeptide would be helpful in a vaccine setting as would generation of cytotoxic T cells. The non-structural region represents such a candidate antigen, but processing by the protease generates several polypeptides, making purification complicated. It would be advantageous, therefore, to derive a non-structural cassette that is unprocessed by the NS3 protease.
- 25 The present invention solves this and other problems using compositions and methods involving an N-terminal deletion in NS3, which removes the catalytic domain. As such, some or all of the remainder of the non-structural region (through NS5B) is expressed as an intact polypeptide. Expression of this species has been documented in mammalian cells as well as in yeast. Further, in certain aspects, polynucleotides
- 30 encoding HCV core polypeptides (or fragments thereof) are added (*e.g.*, operably linked) to the carboxy-terminus of the non-structural cassette. As the core coding region is relatively highly conserved among HCV isolates, the presence of this region

may enhance the immune response. Because core has at its C-terminus a very hydrophobic domain (amino acids 174-191), shorter versions of core were also engineered onto the polypeptide. As described in detail herein, the truncation of core to amino acid 121 yielded higher expression than the amino acid 173 truncation when 5 engineered onto the C-terminus of the mutant NS polypeptide. The combination of most of the non-structural region fused to a C-terminally truncated core into a polypeptide is novel and has advantages for vaccine immunization. Moreover, because the aim is not necessarily to generate antibody responses to this polypeptide, there is no need to maintain a native conformation, enabling a more facile purification protocol.

10

Mutant HCV Non-Structural Polypeptides

Genomes of HCV strains contain a single open reading frame of approximately 9,000 to 12,000 nucleotides, which is transcribed into a polyprotein. An HCV polyprotein is cleaved to produce at least ten distinct products, in the order of NH₂- 15 Core-E1-E2-p7-NS2-NS3-NS4a-NS4b-NS5a-NS5b-COOH. Mutant HCV polypeptides of the invention contain an N-terminal deletion in NS3, which removes or disables the catalytic domain. Preferably, the polypeptides also include the remainder of the non-structural region, although in certain embodiments, the polypeptides may include less than all of the remaining NS polypeptides, for example mutant NS 20 polypeptides including any combinations of NS2-NS3-NS4a-NS4b-NS5a-NS5b (e.g., NS3-NS3-NS5a-NS5b; NS3-NS4a-NS4b; NS3-NS4a-NS4b-NS5a; NS3-NS4b-NS5a-NS5b; NS3-NS4a-NS5a; NS3-NS4b-NS5a; NS3-NS4b-NS5b; etc.).

The HCV NS3 protein functions as a protease and a helicase and occurs at approximately amino acid 1027 to amino acid 1657 of the polyprotein (numbered 25 relative to HCV-1). *See Choo et al. (1991) Proc. Natl. Acad. Sci. USA 88:2451-2455.* HCV NS4 occurs at approximately amino acid 1658 to amino acid 1972, NS5a occurs at approximately amino acid 1973 to amino acid 2420, and HCV NS5b occurs at approximately amino acid 2421 to amino acid 3011 of the polyprotein (numbered relative to HCV-1) (Choo *et al.*, 1991).

30 The mutant polypeptides described herein can either be full-length polypeptides or portions of NS3, NS4 (NS4a and NS4b), NS5a, and NS5b polypeptides. Epitopes of NS3, NS4 (NS4a and NS4b), NS5a, NS5b, NS3NS4NS5a, and NS3NS4NS5aNS5b can

be identified by several methods. For example, NS3, NS4, NS5a, NS5b polypeptides or fusion proteins comprising any combination of the above, can be isolated, for example, by immunoaffinity purification using a monoclonal antibody for the polypeptide or protein. The isolated protein sequence can then be screened by 5 preparing a series of short peptides by proteolytic cleavage of the purified protein, which together span the entire protein sequence. By starting with, for example, 100-mer polypeptides, each polypeptide can be tested for the presence of epitopes recognized by a T cell receptor on an HCV-activated T cell, progressively smaller and overlapping fragments can then be tested from an identified 100-mer to map the epitope 10 of interest.

Epitopes recognized by a T cell receptor on an HCV-activated T cell can be identified by, for example, ^{51}Cr release assay (see Example 2) or by lymphoproliferation assay (see Example 4). In a ^{51}Cr release assay, target cells can be constructed that display the epitope of interest by cloning a polynucleotide encoding the 15 epitope into an expression vector and transforming the expression vector into the target cells. Non-structural polypeptides can occur in any order in the fusion protein. If desired, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more of one or more of the polypeptides may occur in the fusion protein. Multiple viral strains of HCV occur, and NS3, NS4, NS5a, and NS5b polypeptides of any of these strains can be used in a fusion protein.

20 Nucleic acid and amino acid sequences of a number of HCV strains and isolates, including nucleic acid and amino acid sequences of NS3, NS4, NS5a, NS5b genes and polypeptides have been determined. For example, isolate HCV J1.1 is described in Kubo *et al.* (1989) Japan. Nucl. Acids Res. 17:10367-10372; Takeuchi *et al.* (1990) Gene 91:287-291; Takeuchi *et al.* (1990) J. Gen. Virol. 71:3027-3033; and 25 Takeuchi *et al.* (1990) Nucl. Acids Res. 18:4626. The complete coding sequences of two independent isolates, HCV-J and BK, are described by Kato *et al.*, (1990) Proc. Natl. Acad. Sci. USA 87:9524-9528 and Takamizawa *et al.*, (1991) J. Virol. 65:1105-1113 respectively.

Publications that describe HCV-1 isolates include Choo *et al.* (1990) Brit. Med. 30 Bull. 46:423-441; Choo *et al.* (1991) Proc. Natl. Acad. Sci. USA 88:2451-2455 and Han *et al.* (1991) Proc. Natl. Acad. Sci. USA 88:1711-1715. HCV isolates HC-J1 and HC-J4 are described in Okamoto *et al.* (1991) Japan J. Exp. Med. 60:167-177. HCV

isolates HCT 18~, HCT 23, Th, HCT 27, EC1 and EC10 are described in Weiner *et al.* (1991) *Virol.* 180:842-848. HCV isolates Pt-1, HCV-K1 and HCV-K2 are described in Enomoto *et al.* (1990) *Biochem. Biophys. Res. Commun.* 170:1021-1025. HCV isolates A, C, D & E are described in Tsukiyama-Kohara *et al.* (1991) *Virus Genes* 5:243-254.

Each of the mutant HCV polypeptides containing at least portions of NS3, NS4 and NS5 can be obtained from the same HCV strain or isolate or from different HCV strains or isolates. Thus, each non-structural region of the polypeptide can be from the same HCV strain or isolate or from each different HCV strains or isolates. In addition to the mutant HCV non-structural polypeptides described herein, the proteins can contain other polypeptides derived from the HCV polyprotein. For example, it may be desirable to include polypeptides derived from the core region of the HCV polyprotein. This region occurs at amino acid positions 1-191 of the HCV polyprotein, numbered relative to HCV-1. Either the full-length protein or epitopes of the full-length protein may be used in the subject fusions, such as those epitopes found between amino acids 10-53, amino acids 10-45, amino acids 67-88, amino acids 120-130, or any of the core epitopes identified in, e.g., Houghton *et al.*, U.S. Patent No. 5,350,671; Chien *et al.*, *Proc. Natl. Acad. Sci. USA* (1992) 89:10011-10015; Chien *et al.*, *J. Gastroent. Hepatol.* (1993) 8:S33-39; Chien *et al.*, International Publication No. WO 93/00365; Chien, D.Y., International Publication No. WO 94/01778; and commonly owned, U.S. Patent No. 6,150,087. When present, additional non-structural HCV polypeptides such as core can be obtained from the same HCV strain or isolate or from different HCV strains or isolates.

Preferably, the above-described mutant proteins, as well as the individual components of these proteins, are produced recombinantly. A polynucleotide encoding these proteins can be introduced into an expression vector which can be expressed in a suitable expression system. A variety of bacterial, yeast, mammalian, insect and plant expression systems are available in the art and any such expression system can be used. Optionally, a polynucleotide encoding these proteins can be translated in a cell-free translation system. Such methods are well known in the art. The proteins also can be constructed by solid phase protein synthesis.

If desired, the mutant polypeptides, or the individual components of these polypeptides, also can contain other amino acid sequences, such as amino acid linkers or signal sequences, as well as ligands useful in protein purification, such as glutathione-S-transferase and staphylococcal protein A.

5

Polynucleotides

The polynucleotides of the present invention are not necessarily physically derived from the nucleotide sequences shown, but can be generated in any manner, including, for example, chemical synthesis or DNA replication or reverse transcription 10 or transcription. In addition, combinations of regions corresponding to that of the designated sequences can be modified in ways known to the art to be consistent with an intended use.

The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, can be ligated into 15 expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant polypeptides, and a summary of some of the more common control systems and host cell is given below. The polypeptide produced in such host cells is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use.

20 Purification can be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, alkali resolubilization of insoluble protein, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

25 Polynucleotides contain less than an entire HCV genome and can be RNA or single- or double-stranded DNA. Preferably, the polynucleotides are isolated free of other components, such as proteins and lipids. Polynucleotides of the invention can also comprise other nucleotide sequences, such as sequences coding for linkers, signal sequences, or ligands useful in protein purification such as glutathione-S-transferase 30 and staphylococcal protein A.

Polynucleotides encoding mutant HCV non-structural polypeptides can be isolated from a genomic library derived from nucleic acid sequences present in, for

example, the plasma, serum, or liver homogenate of an HCV infected individual or can be synthesized in the laboratory, for example, using an automatic synthesizer. An amplification method such as PCR can be used to amplify polynucleotides from either HCV genomic DNA or cDNA.

5 Further, while the polypeptides that are not NS3, NS4, or NS5 of HCV of the present invention can comprise a substantially complete viral domain, in many applications all that is required is that the polypeptide comprise an antigenic or immunogenic region of the virus. An antigenic region of a polypeptide is generally relatively small-typically 8 to 10 amino acids or less in length. Fragments of as few as 5
10 amino acids can characterize an antigenic region. These segments can correspond to regions of, for example, C, E1, or E2 epitopes. Accordingly, using the cDNAs of C, E1, or E2 as a basis, DNAs encoding short segments of C, E1, or E2 polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical
15 synthesis.

Polynucleotides encoding the polypeptides described herein can comprise coding sequences for these polypeptides which occur naturally or can be artificial sequences which do not occur in nature. These polynucleotides can be ligated to form a coding sequence for the fusion proteins using standard molecular biology techniques.
20 If desired, polynucleotides can be cloned into an expression vector and transformed into, for example, bacterial, yeast, insect, plant or mammalian cells so that the fusion proteins of the invention can be expressed in and isolated from a cell culture.

The expression of polypeptides containing these domains in a variety of recombinant host cells, including, for example, bacteria, yeast, insect, plant and
25 vertebrate cells, give rise to important immunological reagents which can be used for diagnosis, detection, and vaccines.

The general techniques used in extracting the genome from a virus, preparing and probing a cDNA library, sequencing clones, constructing expression vectors, transforming cells, performing immunological assays such as radioimmunoassays and.
30 ELISA assays, for growing cells in culture, and the like are known in the art and laboratory manuals are available describing these techniques. However, as a general

guide, the following sets forth some sources currently available for such procedures, and for materials useful in carrying them out.

Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, *E. coli* is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid containing operons conferring ampicillin and tetracycline resistance, and the various PUC vectors, which also contain sequences conferring antibiotic resistance markers. These markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the Beta-lactamase (penicillinase) and lactose promoter systems (Chang et al. (1977), Nature 198:1056), the tryptophan (trp) promoter system (Goeddel et al. (1980) Nucleic Acid Res. 8:4057), the lambda-derived P[L] promoter and N gene ribosome binding site (Shimatake et al. (1981) Nature 292:128) and the hybrid tac promoter (De Boer et al. (1983) Proc. Natl. Acad. Sci. U.S.A. 292:128) derived from sequences of the trp and lac UV5 promoters. The foregoing systems are particularly compatible with *E. coli*; if desired, other prokaryotic hosts such as strains of *Bacillus* or *Pseudomonas* may be used, with corresponding control sequences.

Eukaryotic hosts include mammalian and yeast cells in culture systems. Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers (1978), Nature 273:113), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art. Vectors suitable for replication in mammalian cells may include viral replicons, or

sequences which insure integration of the appropriate sequences encoding NANBV epitopes into the host genome.

The vaccinia virus system can also be used to express foreign DNA in mammalian cells. To express heterologous genes, the foreign DNA is usually inserted
5 into the thymidine kinase gene of the vaccinia virus and then infected cells can be selected. This procedure is known in the art and further information can be found in these references (Mackett et al. J. Virol. 49: 857-864 (1984) and Chapter 7 in DNA Cloning, Vol. 2, IRL Press).

Yeast expression systems are also known to one of ordinary skill in the art. A
10 yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g., structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a
15 transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

20 Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-
25 phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences (Myanohara et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:1).

In addition, synthetic promoters which do not occur in nature also function as
30 yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197

and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, (Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109).

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *e.g.*, EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*e.g.*, ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*e.g.*, WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a

- leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids
- 5 which direct the secretion of the protein from the cell.
- DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-
- 10 0 060 057).
- A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated
- 15 alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast
- alphafactor. (e.g., see WO 89/02463.)
- 20 Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences,
- 25 such as those coding for glycolytic enzymes.
- Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g., plasmids) capable of stable
- 30 maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle

vectors include YEp24 (Botstein *et al.* (1979) *Gene* 8:17-24), pCI/1 (Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646), and YRp17 (Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157). In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See e.g., Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome (Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245). An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced (Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750). The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1*

allows yeast to grow in the presence of copper ions (Butt *et al.* (1987) *Microbiol. Rev.* 51:351).

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

- Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts:
- 10 *Candida albicans* (Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142), *Candida maltosa* (Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141). *Hansenula polymorpha* (Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302), *Kluyveromyces fragilis* (Das, *et al.* (1984) *J. Bacteriol.* 158:1165), *Kluyveromyces lactis* (De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135), *Pichia guillermondii* (Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141), *Pichia pastoris* (Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555), *Saccharomyces cerevisiae* (Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163), *Schizosaccharomyces pombe* (Beach and Nurse (1981) *Nature* 300:706), and
- 15 *Yarrowia lipolytica* (Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49).

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. (See e.g., Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*; Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*; Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; *Kluyveromyces*; Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos. 4,837,148 and 4,929,555; *Pichia*; Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.*

153:163 *Saccharomyces*; Beach and Nurse (1981) *Nature* 300:706;
Schizosaccharomyces; Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.*
(1985) *Curr. Genet.* 10:49; *Yarrowia*.

Bacterial expression techniques are known in the art. A bacterial promoter is
5 any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g., structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial
10 promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the
15 operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (E. coli) (Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173). Regulated expression
20 may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria:
25 *Bacillus subtilis* (Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541), *Escherichia coli* (Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776,EP-A-0 136 829 and EP-A-0 136 907),
Streptococcus cremoris (Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655);
30 *Streptococcus lividans* (Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655),
Streptomyces lividans (US patent 4,745,056).

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl₂ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the 5 bacterial species to be transformed. (See e.g., Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus, Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949; Campylobacter, Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 10 16:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; Escherichia; Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 15 Lactobacillus; Fiedler *et al.* (1988) *Anal. Biochem* 170:38, Pseudomonas; Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, Staphylococcus, Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 20 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, Streptococcus).

In addition, viral antigens can be expressed in insect cells by the Baculovirus system. A general guide to Baculovirus expression by Summer and Smith is A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures (Texas 25 Agricultural Experiment Station Bulletin No. 1555). To incorporate the heterologous gene into the Baculovirus genome the gene is first cloned into a transfer vector containing some Baculovirus sequences. This transfer vector, when it is cotransfected with wild-type virus into insect cells, will recombine with the wild-type virus. Usually, the transfer vector will be engineered so that the heterologous gene will disrupt the 30 wild-type Baculovirus polyhedron gene. This disruption enables easy selection of the recombinant virus since the cells infected with the recombinant virus will appear phenotypically different from the cells infected with the wild-type virus. The purified

recombinant virus can be used to infect cells to express the heterologous gene. The foreign protein can be secreted into the medium if a signal peptide is linked in frame to the heterologous gene; otherwise, the protein will be bound in the cell lysates. For further information, see Smith et al *Mol. & Cell. Biol.* 3:2156-2165 (1983) or Luckow
5 and Summers in *Virology* 17: 31-39 (1989).

Baculovirus expression can also be affected in plant cells. There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic
10 expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et
15 al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsel et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman
20 Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987).

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*,
25 *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*,

Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

- Transformation can be by any method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and
- 5 transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972), Proc. Natl. Acad. Sci. U.S.A. 69:2110; Maniatis et al. (1982), MOLECULAR CLONING; A LABORATORY MANUAL (Cold Spring Harbor Press,
- 10 Cold Spring Harbor, N.Y.). Yeast transformation by direct uptake may be carried out using the method of Hinnen et al. (1978) Proc. Natl. Acad. Sci. U.S.A. 75: 1929. Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), Virology 52:546 or the various known modifications thereof.
- 15 Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in
- 20 Methods in Enzymology (1980) 65:499-560. Sticky ended cleavage fragments may be blunt ended using E. coli DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.
- 25 Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation
- 30 of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation. Ligation mixtures are transformed into suitable cloning hosts,

such as *E. coli*, and successful transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984), DNA 3:401. If desired, the synthetic strands 5 may be labeled with ^{32}P by treatment with polynucleotide kinase in the presence of ^{32}P -ATP, using standard conditions for the reaction. DNA sequences, including those isolated from cDNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982), Nucleic Acids Res. 10:6487.

10 The expression constructs of the present invention, including the desired fusion, or individual expression constructs comprising the individual components of these fusions, may be used for nucleic acid immunization, to activate HCV-specific T cells, using standard gene delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be 15 delivered either directly to the vertebrate subject or, alternatively, delivered *ex vivo*, to cells derived from the subject and the cells reimplanted in the subject. For example, the constructs can be delivered as plasmid DNA, e.g., contained within a plasmid, such as pBR322, pUC, or ColE1

20 Additionally, the expression constructs can be packaged in liposomes prior to delivery to the cells. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta*. 25 (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

30 Liposomal preparations for use with the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy]propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other

commercially available lipids include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a 5 description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes. The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in METHODS OF IMMUNOLOGY (1983), Vol. 101, pp. 512-527; Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; Papahadjopoulos et al., *Biochim. Biophys. Acta* (1975) 394:483; Wilson et al., *Cell* (1979) 17:77); Deamer and Bangham, *Biochim. Biophys. Acta* (1976) 443:629; Ostro et al., *Biochem. Biophys. Res. Commun.* (1977) 76:836; Fraley et al., *Proc. Natl. Acad. Sci. USA* (1979) 76:3348); Enoch and Strittmatter, *Proc. Natl. Acad. Sci. USA* (1979) 76:145); Fraley et al., *J. Biol. Chem.* (1980) 255:10431; Szoka and Papahadjopoulos, *Proc. Natl. Acad. Sci. USA* (1978) 10 75:145; and Schaefer-Ridder et al., *Science* (1982) 215:166.

The DNA can also be delivered in cochleate lipid compositions similar to those described by Papahadjopoulos et al., *Biochim. Biophys. Acta* (1975) 394:483-491. See, also, U.S. Patent Nos. 4,663,161 and 4,871,488.

A number of viral based systems have been developed for gene transfer into 20 mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems, such as murine sarcoma virus, mouse mammary tumor virus, Moloney murine leukemia virus, and Rous sarcoma virus. A selected gene can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject 25 either *in vivo* or *ex vivo*. A number of retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109. Briefly, retroviral gene delivery 30 vehicles of the present invention may be readily constructed from a wide variety of retroviruses, including for example, B, C, and D type retroviruses as well as spumaviruses and lentiviruses such as FIV, HIV, HIV-1, HIV-2 and SIV (see RNA

Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985). Such retroviruses may be readily obtained from depositories or collections such as the American Type Culture Collection ("ATCC"; 10801 University Blvd., Manassas, VA 20110-2209), or isolated from known sources using commonly available techniques.

5 A number of adenovirus vectors have also been described, such as adenovirus Type 2 and Type 5 vectors. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

10

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

15 Members of the Alphavirus genus, such as but not limited to vectors derived from the Sindbis and Semliki Forest viruses, VEE, will also find use as viral vectors for delivering the gene of interest. For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-
20 519; and International Publication Nos. WO 95/07995 and WO 96/17072.

Other vectors can be used, including but not limited to simian virus 40, cytomegalovirus. Bacterial vectors, such as *Salmonella* ssp. *Yersinia enterocolitica*, *Shigella* spp., *Vibrio cholerae*, *Mycobacterium* strain BCG, and *Listeria monocytogenes* can be used. Minichromosomes such as MC and MC1, bacteriophages, 25 cosmids (plasmids into which phage lambda *cos* sites have been inserted) and replicons (genetic elements that are capable of replication under their own control in a cell) can also be used.

The expression constructs may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected molecule 30 to the immune system and promote trapping and retention of molecules in local lymph nodes. The particles can be phagocytosed by macrophages and can enhance antigen presentation through cytokine release. Examples of particulate carriers include those

derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; and McGee et al., *J. Microencap.* (1996).

- A wide variety of other methods can be used to deliver the expression constructs to cells. Such methods include DEAE dextran-mediated transfection, calcium phosphate precipitation, polylysine- or polyornithine-mediated transfection, or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like. Other useful methods of transfection include electroporation, sonoporation, protoplast fusion, liposomes, peptoid delivery, or microinjection. See, e.g., Sambrook et al., *supra*, for a discussion of techniques for transforming cells of interest; and Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. One particularly effective method of delivering DNA using electroporation is described in International Publication No. WO/0045823.

Additionally, biostatic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering the expression constructs of the present invention. The particles are coated with the construct to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744.

Compositions

The invention also provides compositions comprising the HCV polypeptides or polynucleotides described herein. Such compositions are useful as diagnostics, for example, using the mutant polypeptides (or polynucleotides encoding these polypeptides) in diagnostic reagents. Diagnostics using polypeptides and polynucleotides are known to those of skill in the art.

In addition, immunogenic compounds can be prepared from one or more immunogenic polypeptides derived from the polypeptides described herein, for example the ΔNS35 polypeptide. The preparation of immunogenic compounds which

contain immunogenic polypeptide(s) as active ingredients is known to one skilled in the art. Typically, such immunogenic compounds are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified, or 5 the protein encapsulated in liposomes.

Immunogenic and diagnostic compositions of the invention preferably comprise a pharmaceutically acceptable carrier. The carrier should not itself induce the production of antibodies harmful to the host. Pharmaceutically acceptable carriers are well known to those in the art. Such carriers include, but are not limited to, large, 10 slowly metabolized, macromolecules, such as proteins, polysaccharides such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like, polylactic acids, polyglycolic acids, polymeric amino acids such as polyglutamic acid, polylysine, and the like, amino acid copolymers, and inactive virus particles.

Pharmaceutically acceptable salts can also be used in compositions of the 15 invention, for example, mineral salts such as hydrochlorides, hydrobromides, phosphates, or sulfates, as well as salts of organic acids such as acetates, propionates, malonates, or benzoates. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those of skill in the art. Compositions 20 of the invention can also contain liquids or excipients, such as water, saline, glycerol, dextrose, ethanol, or the like, singly or in combination, as well as substances such as wetting agents, emulsifying agents, or pH buffering agents. Liposomes can also be used as a carrier for a composition of the invention, such liposomes are described above.

25 If desired, co-stimulatory molecules which improve immunogen presentation to lymphocytes, such as B7-1 or B7-2, or cytokines such as GM-CSF, IL-2, and IL-12, can be included in a composition of the invention. Optionally, adjuvants can also be included in a composition. Adjuvants which can be used include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, 30 aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (PCT Publ. No. WO 90/14837),

containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE), formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA),
5 (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton
10 (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (e.g., IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g., gamma
15 interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an E. coli heat-labile toxin (LT), particularly LT-K63, LT-R72, CT-S109, PT-K9/G129; see, e.g., WO 93/13302 and WO 92/19265; (7) other substances that act as immunostimulating agents to enhance
20 the effectiveness of the composition; and (8) microparticles with adsorbed macromolecules, as described in copending U.S. Patent Application Serial No. 09/285,855 (filed April 2, 1999) and international Patent Application Serial No. PCT/US99/17308 (filed July 29, 1999). Alum and MF59 are preferred. The effectiveness of an adjuvant can be determined by measuring the amount of antibodies
25 directed against an immunogenic polypeptide containing an HCV antigenic sequence resulting from administration of this polypeptide in immunogenic compounds which are also comprised of the various adjuvants.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), -acetyl-normuramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), etc.

Thus, such recombinant or synthetic HCV polypeptides can be used in vaccines and as diagnostics. Further, antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy. In addition, antibodies to these polypeptides are useful for isolating and identifying HCV particles.

5 Native HCV antigens can also be isolated from HCV virions. The virions can be grown in HCV infected cells in tissue culture, or in an infected host.

Administration and Delivery

The polynucleotide and polypeptide compositions described herein (e.g., 10 immunogenic compounds) may be administered to a subject using any suitable delivery means. Methods of delivering nucleic acids into host cells are discussed above. Further, HCV polynucleotides and/or polypeptides can be administered parenterally, by injection, usually, subcutaneously, intramuscularly, transdermally or transcutaneously. Certain adjuvants, e.g. LTK63, LTR72 or PLG formulations, can be administered 15 intranasally or orally. Additional formulations which are suitable for other modes of administration include suppositories. For suppositories, traditional binders and carriers can include, for example, polyalkylene glycols or triglycerides; such suppositories can be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Other oral formulations include such normally employed excipients 20 as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.
25 The polypeptides of the present invention can be formulated into the immunogenic compound as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the 30 like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric

hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The immunogenic compounds are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or 5 therapeutically effective. The quantity to be administered, which is generally in the range of 5 micrograms to 250 micrograms of polypeptide per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and can be peculiar to 10 each subject.

The immunogenic compound can be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination can be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune 15 response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. Further, the course of administration may include polynucleotides and polypeptides, together or sequentially (for example, priming with a polynucleotide composition and boosting with a polypeptide composition). The dosage regimen will also, at least in part, be determined by the need of the individual and be 20 dependent upon the judgment of the practitioner.

In certain embodiments, administration of the polynucleotides and polypeptides described herein is used to activate T cells. In addition to the practical advantages of simplicity of construction and modification, administration of polynucleotides encoding mutant NS polypeptides results in the synthesis of a mutant NS polypeptide in the host. 25 Thus, these immunogens are presented to the host immune system with native post-translational modifications, structure, and conformation. The polynucleotides are preferably injected intramuscularly to a large mammal, such as a human, at a dose of 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 5 or 10 mg/kg.

The proteins and/or polynucleotides can be administered either to a mammal 30 which is not infected with an HCV or can be administered to an HCV-infected mammal. The particular dosages of the polynucleotides or fusion proteins in a composition or will depend on many factors including, but not limited to the species,

age, and general condition of the mammal to which the composition is administered, and the mode of administration of the composition. An effective amount of the composition of the invention can be readily determined using only routine experimentation. *In vitro* and *in vivo* models can be employed to identify appropriate doses. Generally, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 5 or 10 mg will be administered to a large mammal, such as a baboon, chimpanzee, or human. If desired, co-stimulatory molecules or adjuvants can also be provided before, after, or together with the compositions.

10 **Antibodies and Diagnostics**

Antibodies, both monoclonal and polyclonal, which are directed against HCV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies.

15 Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. Techniques for raising anti-idiotype antibodies are known in the art. See, e.g., Grzych (1985), Nature 316:74; MacNamara et al. (1984), Science 226:1325, Uytdehaag et al (1985), J. Immunol. 134:1225. These anti-idiotype antibodies may also be useful for treatment
20 and/or diagnosis of NANBH, as well as for an elucidation of the immunogenic regions of HCV antigens.

An immunoassay for viral antigen may use, for example, a monoclonal antibody directed towards a viral epitope, a combination of monoclonal antibodies directed towards epitopes of one viral polypeptide, monoclonal antibodies directed towards 25 epitopes of different viral polypeptides, polyclonal antibodies directed towards the same viral antigen, polyclonal antibodies directed towards different viral antigens or a combination of monoclonal and polyclonal antibodies.

Immunoassay protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, 30 or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide. The labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also

known. Examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

An enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration.

The HCV fusion proteins, such as NS3 mutant and core fusion proteins, can also be used to produce HCV-specific polyclonal and monoclonal antibodies. HCV-specific polyclonal and monoclonal antibodies specifically bind to HCV antigens.

Polyclonal antibodies can be produced by administering the fusion protein to a mammal, such as a mouse, a rabbit, a goat, or a horse. Serum from the immunized animal is collected and the antibodies are purified from the plasma by, for example, precipitation with ammonium sulfate, followed by chromatography, preferably affinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art.

Monoclonal antibodies directed against HCV-specific epitopes present in the fusion proteins can also be readily produced. Normal B cells from a mammal, such as a mouse, immunized with, e.g., a mutant NS3 polypeptide or NS-core fusion protein can be fused with, for example, HAT-sensitive mouse myeloma cells to produce hybridomas. Hybridomas producing HCV-specific antibodies can be identified using

RIA or ELISA and isolated by cloning in semi-solid agar or by limiting dilution.

Clones producing HCV-specific antibodies are isolated by another round of screening.

Antibodies, either monoclonal and polyclonal, which are directed against HCV epitopes, are particularly useful for detecting the presence of HCV or HCV antigens in a sample, such as a serum sample from an HCV-infected human. An immunoassay for an HCV antigen may utilize one antibody or several antibodies. An immunoassay for an HCV antigen may use, for example, a monoclonal antibody directed towards an HCV epitope, a combination of monoclonal antibodies directed towards epitopes of one HCV polypeptide, monoclonal antibodies directed towards epitopes of different HCV polypeptides, polyclonal antibodies directed towards the same HCV antigen, polyclonal antibodies directed towards different HCV antigens, or a combination of monoclonal and polyclonal antibodies. Immunoassay protocols may be based, for example, upon competition, direct reaction, or sandwich type assays using, for example, labeled antibody. The labels may be, for example, fluorescent, chemiluminescent, or radioactive.

The polyclonal or monoclonal antibodies may further be used to isolate HCV particles or antigens by immunoaffinity columns. The antibodies can be affixed to a solid support by, for example, adsorption or by covalent linkage so that the antibodies retain their immunoselective activity. Optionally, spacer groups may be included so that the antigen binding site of the antibody remains accessible. The immobilized antibodies can then be used to bind HCV particles or antigens from a biological sample, such as blood or plasma. The bound HCV particles or antigens are recovered from the column matrix by, for example, a change in pH.

25 **Methods of Eliciting Immune Responses**

HCV-specific T cells that are activated by the above-described polypeptides, expressed *in vivo* or *in vitro* preferably recognize an epitope of an HCV polypeptide such as a mutant NS3 polypeptide, including an epitope of a mutant HCV polypeptide. HCV-specific T cells can be CD8⁺ or CD4⁺.

30 HCV-specific CD8⁺ T cells preferably are cytotoxic T lymphocytes (CTL) which can kill HCV-infected cells that display NS3, NS4, NS5a, NS5b epitopes complexed with an MHC class I molecule. HCV-specific CD8⁺ T cells may also

express interferon- γ (IFN- γ). HCV-specific CD8 $^{+}$ T cells can be detected by, for example, ^{51}Cr release assays. ^{51}Cr release assays measure the ability of HCV-specific CD8 $^{+}$ T cells to lyse target cells displaying an nonstructural (*e.g.*, mutant NS) epitope. HCV-specific CD8 $^{+}$ T cells which express IFN- γ can also be detected by

5 immunological methods, preferably by intracellular staining for IFN- γ after *in vitro* stimulation with a mutant NS polypeptide.

HCV-specific CD4 $^{+}$ cells activated by the above-described polypeptides, expressed *in vivo* or *in vitro*, and combinations of the individual components of these proteins, preferably recognize an epitope of a mutant non-structural polypeptide,

10 including an epitope of a mutant protein, that is bound to an MHC class II molecule on an HCV-infected cell and proliferate in response to stimulating mutant peptides.

HCV-specific CD4 $^{+}$ T cells can be detected by a lymphoproliferation assay. Lymphoproliferation assays measure the ability of HCV-specific CD4 $^{+}$ T cells to proliferate in response to an epitope.

15 Mutant NS (or fusions thereof with core, envelope or other viral polypeptides) can be used to activate HCV-specific T cells either *in vitro* or *in vivo*. Activation of HCV-specific T cells can be used, *inter alia*, to provide model systems to optimize CTL responses to HCV and to provide prophylactic or therapeutic treatment against HCV infection. For *in vitro* activation, proteins are preferably supplied to T cells via a

20 plasmid or a viral vector, such as an adenovirus vector, as described above.

Polyclonal populations of T cells can be derived from the blood, and preferably from peripheral lymphoid organs, such as lymph nodes, spleen, or thymus, of mammals that have been infected with an HCV. Preferred mammals include mice, chimpanzees, baboons, and humans. The HCV serves to expand the number of activated HCV-specific T cells in the mammal. The HCV-specific T cells derived from the mammal can then be restimulated *in vitro* by adding HCV epitopic peptides to the T cells. The HCV-specific T cells can then be tested for, *inter alia*, proliferation (*e.g.*, lymphoproliferation assays known in the art), the production of IFN- γ , and the ability to lyse target cells displaying HCV NS epitopes *in vitro*.

The following examples are meant to illustrate the invention and are not meant to limit it in any way. Those of ordinary skill in the art will recognize modifications within the spirit and scope of the invention as set forth herein.

5

EXAMPLES

Example 1: Constructs

pCMV-II: pCMV-II (Figure 7, SEQ ID NO:5) was created to contain the human CMV promoter, enhancer, intron A, polylinker and the bovine growth hormone 10 terminator in a deleted-pUC backbone (Life Technologies).

pT7-HCV: pT7-HCV was created in a polylinker-modified pUC vector to contain full-length HCV cDNA preceded by a synthetic T7 promoter. pT7-HCV also contains the complete 5' UTR and the poly A version of the 3' UTR.

pCMV.ΔNS35: To generate pCMV.ΔNS35 (Figure 5, SEQ ID NO:3), a two 15 step procedure was undertaken. First, a PCR product was generated from pT7-HCV that corresponded to the following: a 5' EcoRI site, followed by the Kozak sequence of ACCATGG; the initiator ATG followed by amino acid #1242 and continuing to the StuI site. Second, the StuI to XbaI fragment from a full-length genomic clone was isolated. The genomic clone consisted of the T7 promoter fused to the full-length HCV 20 cDNA with the poly A version of the 3' end, in a pUC vector. Finally, the EcoRI-StuI and StuI-XbaI fragments were ligated into the pCMV-II expression vector, transformed into HB101 competent cells and plated onto ampicillin (100 µg/ml). Miniprep analyses led to the identification of the desired clone which was amplified on a larger scale using a Quigen Gigaprep kit following the manufacturer's specifications. The resulting clone 25 was named pCMV.ΔNS35 (Figure 5, SEQ ID NO:3).

pd.ΔNS3NS5: As shown schematically in Figure 10, the yeast expression plasmid pd.ΔNS3NS5 (SEQ ID NO:8) was constructed using restriction fragments obtained from the mammalian expression plasmid pCMV.KM.ΔNS35. pCMV.KM.ΔNS35 is identical to pCMV.ΔNS35 (Figure 5, SEQ ID NO:3) except that 30 it contains a kanamycin resistance gene in the viral backbone. pCMV.KM.ΔNS35 was digested with EcoRI and NheI to obtain 2895bp EcoRI-NheI fragment. EcoRI-NheI

fragment was ligated into pRSET HindIII-NheI subcloning vector with oligos (HE) from HindIII to EcoRI. After sequence verification, pRSETHindIII-NheI #6 was digested with HindIII and NheI to obtain a 2908bp HindIII-NheI fragment.

pCMV.KM. Δ NS35 was linearized with XbaI and ligated with synthetic oligos (XS) from XbaI-SalI. The ligation was digested with NheI and SalI to obtain 2481bp NheI-SalI fragment. The fragment was ligated into pET3a NheI-SalI subcloning vector. After sequence verification, pET3a NheI-SalI #2 was digested with NheI and SalI to obtain a 2481bp NheI-SalI fragment. BamHI-HindIII ADH2/GAPDH promoter fragment was then ligated with HindIII-NheI and NheI-SalI fragments into pBS24.1 10 BamHI-SalI yeast expression vector.

pd. Δ NS3NS5.PJ: pd. Δ NS3NS5.PJ (Figures 13 and 14; SEQ ID NO:10) was generated to create a "perfect junction" at the 5' and 3' end of the HCV coding region. At the 5' end of pd. Δ NS3NS5, there were 6 extra bases between the yeast ADH2/GAPDH promoter and the ATG of the polypeptide. At the 3' end, there were 52 15 bases of untranslated sequence between the stop codon of the polypeptide and the α -factor terminator in the yeast expression vector. pd. Δ NS3NS5.PJ was created by digesting pd. Δ NS3NS5 #17 with ScaI and SphI to obtain 4963bp ScaI-SphI fragment. pd.NS5b3011 was digested with SphI and SalI to obtain a 321bp SphI-SalI fragment which gave the "perfect junction" at the 3' end of the polypeptide. The ScaI-SphI and 20 SphI-SalI fragments were ligated into pSP72 HindIII-SalI subcloning vector with synthetic oligos from HindIII-ScaI(HS) for the "perfect junction" at the 5' end.

The region of synthetic sequence in pSP72 HindIII-SalI clone# 6 was verified. pSP72 HindIII-SalI clone#6 was digested with HindIII and BlnI or with BlnI and SalI to obtain 2441bp HindIII-BlnI and 2895bp BlnI-SalI fragments, respectively. The 25 BamHI-HindIII ADH2/GAPDH promoter fragment was ligated to HindIII-BlnI and BlnI-SalI fragments into pBS24.1 BamHI-SalI yeast expression vector.

pd. Δ NS3NS5.PJ.core121RT and pd. Δ NS3NS5.PJ.core173RT were generated and encode HCV core aa 1-121 at the C-terminus of the Δ NS3NS5 polypeptide (designated pd. Δ NS3NS5.PJ.core121RT, SEQ ID NO:12) and core aa 1-173 at the C-terminus of the Δ NS3NS5 polypeptide (designated pd. Δ NS3NS5.PJ.core173RT, SEQ 30 ID NO:14). The core sequence had aa 9 mutated from Lys to Arg and aa 11 mutated

from Asn to Thr, designated as core 121RT or 173RT.

- pd.ΔNS3NS5.PJ.core121RT and pd.ΔNS3NS5.PJ.core173RT: To generate pd.ΔNS3NS5.PJ.core121RT (Figure 17, SEQ ID NO:12) and pd.ΔNS3NS5.PJ.core173RT (Figure 18, SEQ ID NO:14). As shown in Figure 16, a
- 5 NotI-Sal HCVcore121RT and HCVcore173RT were amplified by PCR, from an *E. coli* expression plasmid, pSODCF2.HCVcore191RT #2. Either the core 121RT Not-SalI PCR product or the core 173RT Not-SalI PCR product were ligated into a pT7Blue2 PstI-SalI subcloning vector with synthetic oligos (PN) from PstI to NotI. After sequence confirmation, pT7Blue2core121RT clone#9 and pT7Blue2core173RT
- 10 clone#11 was digested with PstI and SalI to obtain 403bp and 559bp PstI-SalI fragments, respectively, for further cloning.

A 121bp NotI-PstI fragment from pSP72 HindIII-SalI clone #6 was isolated as described above during the cloning of pd.ΔNS3NS5.PJ. NotI-PstI and PstI-SalI fragments were assembled into a vector made by digesting pd.ΔNS3NS5.PJ clone#5

15 (described above) with NotI and SalI.

ΔNS3NS5 and Core 140 and Core 150: An HCV core epitope was found which elicits CTLs in baboons (HCV core aa 121-135). Since pd.ΔNS3NS5.PJ.core121RT ends right before this potentially important epitope and was expressed better than the longer pd.ΔNS3NS5.PJ.core173RT construct (Example 2), two intermediate constructs

20 were made which include this epitope, possibly giving intermediate expression levels. The two new constructs fused HCV core aa 1-140 or HCV core aa1-150 to the C terminus of ΔNS3NS5.PJ.

pd.ΔNS3NS5.PJ.core140RT (Figure 21, SEQ ID NO:16) and
pd.ΔNS3NS5.PJ.core150RT (Figure 22, SEQ ID NO:18): As shown in Figure 20, a

25 PstI-SalI HCVcore140RT and a PstI-SalIHCVcore150RT fragment were amplified by PCR from pd.ΔNS3NS5.PJ.core173RT clone #16. Ligate either HCV core PstI-SalI PCR products into pT7Blue2 PstI-SalI subcloning vector. After sequence confirmation, pT7Blue2core140RT clone#22 and pT7Blue2core150RT clone#26 were digested with PstI-SalI to obtain 460bp and 490bp PstI-SalI fragments, respectively, for

30 further cloning.

A 121bp NotI-PstI fragment was isolated from pSP72 HindIII-SalI clone #6 (as described above during the cloning of pd.ΔNS3NS5.PJ. NotI-PstI and PstI-SalI fragments were assembled into a vector made by digesting pd.ΔNS3NS5.PJ clone#5 (described above) with NotI and SalI.

5

Example 2: Protein Expression

Various of the constructs described herein, encoding HCV-1 ΔNS3 to NS5 antigen (aa 1242-3011), were expressed in yeast. *S. cerevisiae* strain AD3 was transformed with pd.ΔNS3NS5 and checked for expression. A stained protein band at 10 the expected molecular weight of 194 kD was not observed (Figure 12). Strain AD3 was also transformed with pd.ΔNS3NS5.PJ clone #5 and checked for expression. A protein band of the expected molecular weight of 194kD was detected (Figure 15).

Strain AD3 was transformed with pd.ΔNS3NS5.PJ.core121RT clone #6 and pd.ΔNS3NS5.PJ.core173RT clone#15 and checked for expression. Protein bands of the 15 expected molecular weight of 206kD and 210kD, respectively, were observed. Expression levels of the pd.ΔNS3NS5.PJ.core173RT construct were much less than that of the pd.ΔNS3NS5.PJ.core121RT construct. (See Figure19). Thus, there is a correlation of protein expression levels and the length of HCV core.

Strain AD3 were transformed with pd.ΔNS3NS5.PJ.core140RT clone# 29 and 20 pd.ΔNS3NS5.PJ.core150RT clone#35 and checked for expression. Bands of the expected molecular weights of 208kD and 209kD were seen by stain at levels close to those of pd.ΔNS3NS5core173RT (Figure 23).

Example 3: Eliciting Immune Responses

A. Immunization

To evaluate the immunogenicity of the mutant NS polypeptides, studies using 5 guinea pigs, rabbits, mice, rhesus macaques and/or baboons are performed. The studies are structured as follows: DNA immunization alone (single or multiple); DNA immunization followed by protein immunization (boost); DNA immunization followed 30 by protein immunization; immunization by PLG particles. Immunization is intramuscular or mucosally.

B. Humoral Immune Response

The humoral immune response is checked in serum specimens from immunized animals with anti-NS antibody ELISAs (enzyme-linked immunosorbent assays) at various times post-immunization. Briefly, serum from immunized animals is screened for antibodies directed against the NS or mutant NS proteins. Wells of ELISA microtiter plates are coated overnight with the selected HCV protein and washed four times; subsequently, blocking is done with PBS-0.2% Tween (Sigma). After removal of the blocking solution, diluted mouse serum is added. Sera are tested at various dilutions. Microtiter plates are washed and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates are washed and 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) is added per well. The optical density of each well is measured. Titers are typically reported as the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.). Similarly, generation of neutralization of binding (NOB) antibodies can be measured by methods known in the art.

C. Cellular Immune Response

The frequency of specific cytotoxic T-lymphocytes (CTL) is evaluated by a standard chromium release assay of peptide pulsed Balb/c mouse CD4 cells. Briefly, spleen cells (Effector cells, E) are obtained from the BALB/c mice immunized, cultured, restimulated, and assayed for CTL activity against HCV peptide-pulsed target cells. Cytotoxic activity is measured in a standard ⁵¹Cr release assay.

25 Example 4: Immunization with PLG-delivered DNA.

The polylactide-co-glycolide (PLG) polymers are obtained from Boehringer Ingelheim, U.S.A. The PLG polymer is RG505, which has a copolymer ratio of 50/50 and a molecular weight of 65 kDa (manufacturers data). Cationic microparticles with adsorbed DNA are prepared using a modified solvent evaporation process, essentially as described in Singh et al., *Proc. Natl. Acad. Sci. USA* (2000) 97:811-816. Briefly, the microparticles are prepared by emulsifying a 5% w/v polymer solution in methylene chloride with PBS at high speed using an IKA homogenizer. The primary emulsion is

- then added to distilled water containing cetyl trimethyl ammonium bromide (CTAB) (0.5% w/v). This results in the formation of a w/o/w emulsion which was stirred at room temperature, allowing the methylene chloride to evaporate. The resulting microparticles are washed in distilled water by centrifugation and freeze dried.
- 5 Following preparation, washing and collection, DNA is adsorbed onto the microparticles by incubating cationic microparticles in a solution of DNA. The microparticles are then separated by centrifugation, the pellet washed with TE buffer and the microparticles are freeze dried, resuspended and administered to animals. Antibody titers are measured by ELISA assays.

10

What is claimed is:

1. An isolated mutant non-structural ("NS") HCV polypeptide comprising a polypeptide having a mutation in the catalytic domain of NS3, wherein said mutation 5 functionally disrupts the catalytic domain.
2. The polypeptide of claim 1, wherein the mutation comprises a deletion.
3. The polypeptide of claim 1, wherein the mutation comprises a 10 substitution.
4. The polypeptide of any of claims 1-3, wherein said NS polypeptide comprises NS3, NS4 and NS5.
- 15 5. The polypeptide of any of claims 1-3, wherein said NS polypeptide consists of NS3, NS4 and NS5.
6. The polypeptide of any of claims 1-3, wherein said NS polypeptide 20 consists of NS3 and NS5.
7. The polypeptide of claim 6, wherein NS5 consists of NS5a.
8. The polypeptide of claim 6, wherein NS5 consists of NS5b.
- 25 9. The polypeptide of any of claims 1-3, wherein said NS polypeptide consists of NS3 and NS4.
10. The polypeptide of claim 9, wherein NS4 consists of NS4a.
- 30 11. The polypeptide of claim 9, wherein NS4 consists of NS4b.

12. The polypeptide of claim 4, further comprising a second viral polypeptide that is not NS3, NS4, or NS5 of HCV.
13. The polypeptide of claim 12, wherein the second viral polypeptide comprises an HCV Core polypeptide ("C"), or fragment thereof.
14. The polypeptide of claim 13, wherein the C polypeptide is truncated.
15. The polypeptide of claim 14, wherein the truncation is at amino acid 10 121.
16. The polypeptide of claim 12, wherein the polypeptide further comprises an HCV envelope protein ("E").
- 15 17. The polypeptide of claim 16, wherein the E is E1.
18. The polypeptide of claim 16, wherein the E is E2.
19. A composition comprising
20 (a) the polypeptide of any one of claims 1-18; and
(b) a pharmaceutically acceptable excipient.
20. An isolated and purified polynucleotide which encodes the mutant HCV polypeptide according to any one of claims 1-18.
25
21. A composition comprising
(a) the isolated purified polynucleotide of claim 20; and
(b) a pharmaceutically acceptable excipient.
- 30 22. The composition of claim 21, wherein the polynucleotide is DNA.

23. The composition of claim 21, wherein the polynucleotide is in a plasmid.
24. An expression vector comprising the polynucleotide of claim 20.
5
25. An expression vector comprising the polynucleotide of SEQ ID NO:8.
26. A host cell comprising the polynucleotide of claim 20.
10
27. The host cell of claim 26, wherein the cell is a yeast cell.
28. The host cell of claim 26, wherein the cell is a mammalian cell.
29. The host cell of claim 26, wherein the cell is an insect cell.
15
30. The host cell of claim 26, wherein the cell is a plant cell.
31. The host cell of claim 26, wherein the polynucleotide comprises the sequence of SEQ ID NO:8.
20
32. The polypeptide of claim 1, wherein the polypeptide further comprises SEQ ID NO:9.
25
33. A method of preparing a mutant NS HCV polypeptide, wherein the method comprises the steps of:
 - a. transforming a host cell with an expression vector according to claim 24, under conditions wherein the polypeptide is expressed; and
 - b. isolating the polypeptide.
30

34. The method of claim 33, wherein the host cell is a yeast cell.
35. The method of claim 33, wherein the host cell is a mammalian cell.
- 5 36. The method of claim 33, wherein the host cell is an insect cell.
37. The method of claim 33, wherein the host cell is a plant cell.
- 10 38. An antibody that specifically binds to a polypeptide of any of claims 1-18.
39. The antibody of claim 38, wherein the antibody is a monoclonal antibody.
- 15 40. The antibody of claim 38, wherein the antibody is a purified polyclonal antibody.
41. A method of eliciting an immune response in a subject, comprising the step of administering to the subject a polypeptide of any of claims 1-18.
- 20 42. A method of eliciting an immune response in a subject, comprising the step of administering to the subject a polynucleotide of claim 20.

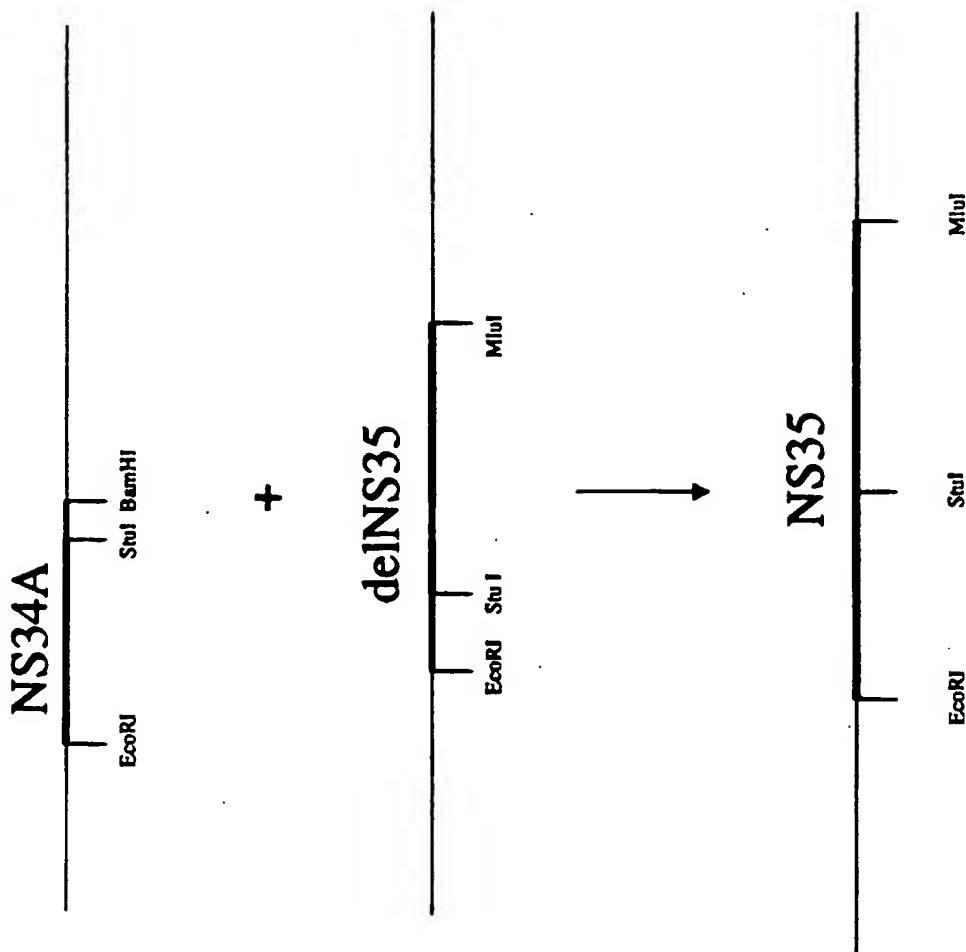
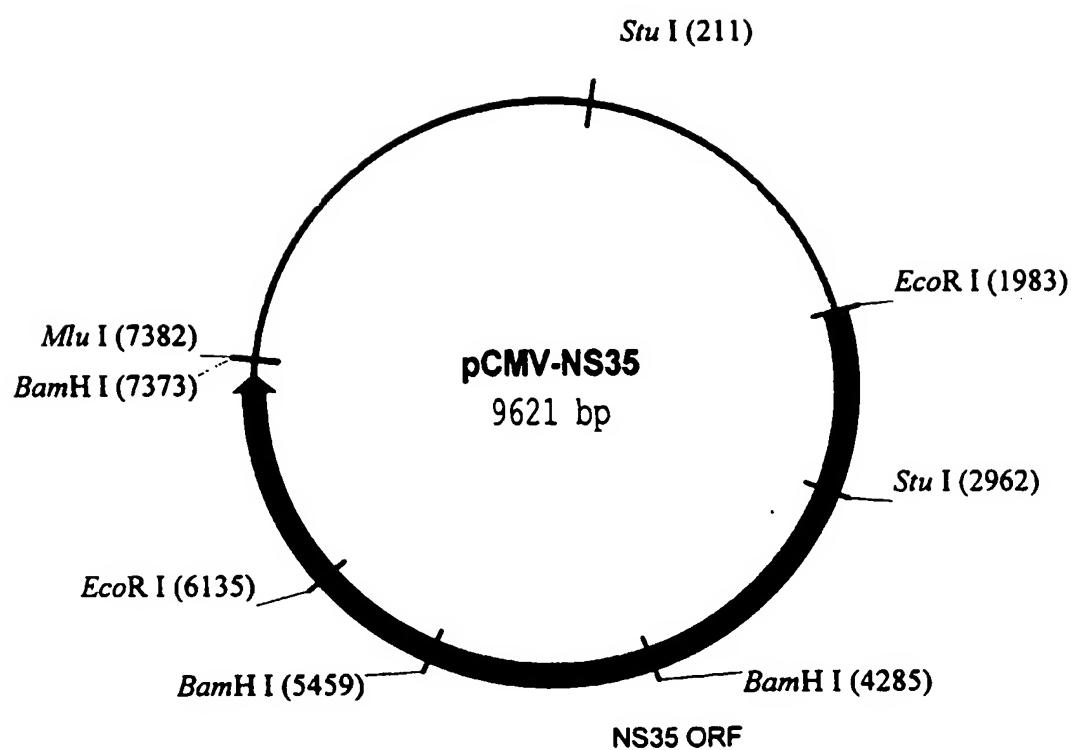
Cloning Scheme for Generating pCMV-NS35

FIGURE 2

pCMV-NS35 9/100

FIGURE 3 - Page 1

1 TCGCGGTTT CGGTGATGAC GGTGAAACCC TCTGACACAT GCAGCTCCCC GAGACGGTCA CAGCTTGCT GTAAAGCGGAT
AGGCCCAAA GCCACTACTG CCACCTTTGG AGACTGTGTA CGTCGAGGGC CTCTGCCAGT GTCGAACAGA CATTGCCTA

81 GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGGGGTG TTGGCGGGTG TCAGGGCTGG CTTAACTATG CGGCATCAGA
CGGCCCTCGT CTGTCGGGC AGTCCCAGC AGTCGCCAC AACCGCCAC AGCCCCGACC GAATTGATAAC GCCGTAGTCT

StuI

161 GCAGATTGTA CTGAGAGTGC ACCATATGAA GCTTTTGCA AAAGCCTAGG CCTCCAAAAA AGCCTCCTCA CTACTTCTGG
CGTCTAACAT GACTCTCACG TGGTATACTT CGAAAACGT TTTCGATCC GGAGGTTTT TCGGAGGAGT GATGAAGACC

241 AATAGCTAG AGGGCGAGGC GGCCCTCGGCC TCTGCATAAA TAAAAAAAT TAGTCAGCCA TGGGGCGGAG AATGGGCGGA
TTATCGAGTC TCCGGCTCCG CCGGAGCCGG AGACGTATT ATTTCCTTA ATCAGTCGGT ACCCCGCCTC TTACCCGCCT

321 ACTGGGCGGG GAGGGAATTG TTGGCTATTG GCCATTCGAT ACCTGATATC TATATCATAA TATGTACATT TATATGGCT
TGACCCGCC CTCCTTAAT AACCGATAAC CGGTAACGTA TGCAACATAG ATATAGTATT ATACATGTAATATAACCGA

401 CATGTCCAAT ATGACCGCA TGTGACATT GATTATTGAC TAGTTATTA TAGTAATCAA TTACGGGTC ATTAGTTCAT
GTACAGGTTA TACTGGGGT ACAACTGTA CTAATAACTG ATCAATAATT ATCATAGT ATTAGCCCAG TAATCAAGTA

481 AGCCCATATA TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCGCC TGGCTGACCG CCCAACGACC CCCGCCATT
TCGGGTATAT ACCTCAAGGC GCAATGTATT GAATGCCATT TACCGGGCGG ACCGACTGGC GGGTGTGG GGGCGGGTAA

561 GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA GGGACTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT
CTGCAGTTAT TACTGCATAC AAGGGTATCA TTGCGGTTAT CCCTGAAAGG TAACTGCAGT TACCCACCTC ATAAATGCCA

641 AACTGCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTCCGCC CCTATTGACG TCAATGACGG TAAATGGCC
TTTACGGGT GAACCGTCAT GTAGTTACA TACTATACGG TTCAGGGGG GGATAACTGC ACTTACTGCC ATTACCGGG

721 GCCTGGCATT ATGCCCAAGTA CATGACCTTA CGGGACTTTC CTACTGGCA GTACATCTAC GTATTAGTC TCGCTATTAC
CGGACCGTAA TACGGGTCA GTACTGGAAT GCCCTGAAAG GATGAACCGT CATGTAGATG CATAATCAGT AGCGATAATG

801 CATGGTGTAG CGGTTTTGGC AGTACACCAA TGGCGTGGA TAGCGTTTG ACTCACGGG ATTCCAAGT CTCCACCCCA
GTACCACTAC GCCAAAACCG TCATGTGGTT ACCGGCACCT ATCGCCAAAC TGAGTCCCC TAAAGTTCA GAGGTGGGGT

881 TTGACGTCAA TGGAGTTTG TTTGGCACC AAAATCAACG GGACTTCCA AAATGCGTA ATAACCCCGC CCCGTTGACG
AACTCCAGTT ACCCTCAAAC AAAACCGTGG TTTAGTTGC CCTGAAAGGT TTACAGCAT TATTGGGGCG GGGCAACTG

961 CAAATGGCG GTAGGCGTGT ACAGTGGGAG GTCTATATAA GCAGAGCTCG TTATGACAC CGTCAGATCG CCTGGAGACG
GTTTACCCGC CATCCGCACA TGCCACCCCTC CAGATATATT CGTCTCGAGC AAATCACTTG GCAGTCTAGC GGACCTCTGC

1041 CCATCCACGC TGTGTTGACCC TCCATAGAAAG ACACCGGGAC CGATCCAGCC TCCGGGGCCG GGAACGGTGC ATTGGAACGC
GGTAGGTGGC ACAAAACTGG AGGTATCTTC TGTGGCCCTG GCTAGGTGGC AGGGCCGGC CCTGCCACG TAACCTTGCG

1121 GGATCCCCG TGCCAAGAGT GACGTAAGTA CGCCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA
CCTAAGGGC ACAGGTTCTCA CTGCTATTCA CGCAGATATC TGAGATATCC GTGTGGGGAA ACCGAGAATA CGTACGATAT

1201 CTGTTTTGG CTTGGGGCCT ATACACCCCG GCTCCTTATG CTATAGGTGA TGGTATAGCT TAGCCTATAG GTGTGGGGTAA
GACAAAAACC GAACCCCGGA TATGTGGGGG CGAGGAATAC GATATCCACT ACCATATCGA ATCGGATATC CACACCAAT

1281 TTGACCATTA TTGACCACTC CCCTATTGGT GACGATACTT TCCATTACTA ATCCATAACA TGGCTCTTG CCACAACTAT
AACTGGTAAT AACTGGTGAG GGGATAACCA CTGCTATGAA AGGTAAATGAT TAGTATTGT ACCGAGAAC GGTGTTGATA

1361 CTCTATTGGC TATATGCCAA TACTCTGTCC TTCAGAGACT GACACGGACT CTGTATTTT ACAGGATGGG GTCCATTAT
GAGATAACCG ATATACGGT ATGAGACAGG AAGTCTGTA CTGTGCCCTGA GACATAAAA TGCTCCTACCC CAGGTAATA

pCMV-NS35

FIGURE 3 - Page 2

1441 TATTTACAAA TTCACATATA CAACAACGCC GTCCCCCGTG CCCGAGTTT TTATTAACAA TAGCGTGGGA TCTCCGACAT
ATAAAATGTTT AAGTGTATAT GTTGTGCGG CAGGGGGCAC GGGCGTCAAA ATAATTGT ATCCGACCCCT AGAGGCTGTA

1521 CTCGGGTACG TGTTCCGGAC ATGGGCTTT CTCCGGTAGC GGCGGAGCTT CCACATCCGA GCCCTGGTCC CATCCGTC
GAGCCCATGC ACAAGGCCTG TACCCGAGAA GAGGCCATCG CGGCGCTGAA GTGTAGGCT CGGGACCAGG GTAGGCAGGT

1601 GCGGCGTCATG GTCGCTCGGC AGCTCCTTC TCCTAACAGT GGAGGCCAGA CTTAGGCACA GCACAATGCC CACCACCA
CGCCGACTAC CAGCGAGCCG TCGAGGAACG AGGATTGTCA CCTCCGGTCT GAATCCGTGT CGTGTACGG GTGGTGGTGG

1681 AGTGTGCCGC ACAAGGCCGT GGCGGTAGGG TATGTGTCTG AAAATGAGCT CGGAGATTGG GCTCCACCT GGACGCAGAT
TCACACGCCG TGTTCCGGCA CCCGCATCCC ATACACAGAC TTTACTCGA GCCTCTAACCG CAGCGTGGGA CCTGCGTCTA

1761 GGAAGACTTA AGGCAGCCGC AGAAGAAGAT GCAGGCAGCT GAGTTGTGT ATTCTGATAA GAGTCAGAGG TRACTCCCGT
CCTTCTGAAT TCGTCGCCG TCTTCTTCTA CGTCCGTCGA CTCAACAAACA TAAGACTTATT CTCACTCTCC ATTGAGGGCA

1841 TGCGGTGCTG TTAACGGTGG AGGGCAGTGT AGTCTGAGCA GTACTCGTGT CTGCCGCCGC CGCCACCAAGA CATAATAGCT
ACGCCACGAC AATTGCCACC TCCCGTCACA TCAGACTCGT CATGAGCAAC GACGGCGCGC GCGGIGGTCT GTATTATCGA

+2 M A A

1921 GACAGACTAA CAGACTGTTG CTTCCATGG GTCTTTCTG CAGTCACCGT CGTCGACCTA AGAATTCAAC ATGGCTGCAT
CTGTCGATT GTCTGACAAG GAAAGGTACC CAGAAAAGAC GTCACTGGCA GCAGCTGGAT TCTTAAGTGG TACCGACGTA

+2 Y A A Q G Y K V L V L N P S V A A T L G F G A Y M S K
2001 ATGCAGCTCA GGGCTATAAG GTGCTAGTAC TCAACCCCTC TGTTGCTGCA ACACCTGGCT TTGGTGCTTA CATGTCAG
TACGTCGAGT CCCGATATTG CACGATCATG AGTTGGGAG ACAACGACGT TGTGACCCGA ACCACGAAT GTACAGGTC

+2 A H G I D P N I R T G V R T I T T G S P I T Y S T Y G
2081 GCTCATGGGA TCGATCTAA CATCAGGACC GGGGTGAGAA CAATTACAC TGGCAGCCCC ATCACGTAAT CCACCTACGG
CGAGTACCT AGCTAGGATT GTAGTCCCTGG CCCCACTCTT GTTAATGGTGT ACCGTCGGGG TAGTGCATGA GGTGGATGCC

+2 K F L A D G G C S G G A Y D I I I C D E C H S T D A
2161 CAAGTTCTT GCCGACGGCG GGTGCTCGG GGGCGTTAT GACATAATAA TTGTGACGA GTGCCACTCC ACCGATGCCA
GTTCAAGGAA CGGCTGCCGC CCACGAGCCC CCCGCGAATA CTGTATTATT AAACACTGCT CACGGTGAGG TGCCTACGGT

+2 T S I L G I G T V L D Q A E T A G A R L V V L A T A T
2241 CATCCATCTT GGGCATTGGC ACTGTCCCTG ACCAAGCAGA GACTGCCGGG GCGAGACTGG TTGTGCTCGC CACCGCCACC
GTAGGTAGAA CCCGTAACCG TGACAGGAAC TGTTGCTCT CTGACCCCC CGCTCTGACC AACACGAGCG GTGGCGGTGG

+2 P P G S V T V P H P N I E E V A L S T T G E I P F Y G
2321 CCTCCGGGCT CCGTCACTGT GCCCCATCCC AACATCGAGG AGGTTGCTCT GTCCACCACC GGAGAGATCC CTTTTACGG
GGAGGCCGA GGCAGTGACA CGGGTAGGG TTGTAGCTC TCCAACGAGA CAGGTGGTGG CCTCTCTAGG GAAAATGCC

+2 K A I P L E V I K G G R H L I F C H S K K K C D E L
2401 CAAGGTATC CCCCTCGAAG TAATCAAGGG GGGGAGACAT CTCATCTCT GTCACTCAA GAAGAAGTGC GACGAACCTCG
GTTCCGATAG GGGGAGCTTC ATTAGTCTCC CCCCTCTGTA GAGTAGAAGA CAGTAAGTTT CTTCTCACG CTGCTTGAGC

+2 A A K L V A L G I N A V A Y Y R G L D V S V I P T S G
2481 CCGCAAGCT GGTGCGATTG GGCATCAATG CGGTGGCTA CTACCGGGT CTTGACGTGT CGGTACCTCC GACCAGCGGC
GGCGTTTGA CCAGCTAAC CGCTAGTTAC GGCACCGGAT GATGGCGCA GAACGACCA GGCAGTAGGG CTGGTCGCCG

+2 D V V V V A T D A L M T G Y T G D F D S V I D C N T C
2561 GATGTTGTCG TCGTGGCAAC CGATGCCCTC ATGACCGGCT ATACCGGCA CTTCGACTCG GTGATAGACT GCAATACGTG
CTACAACAGC AGCACCGTTG GCTACGGGAG TACTGGCCG TATGGCCG TAAAGCTGAGC CACTATCTGA CGTTATGCA

pCMV-NS35

FIGURE 3 - Page 3

+2 V T Q T V D F S L D P T E T I E T I T L P Q D A V S
 2641 . TGTCACCCAG ACAGTCGATT TCAGCCTGAA CCCTACCTTC ACCATGAGA CAATCACGCT CCCCCAAGAT GCTGTCTCCC
 ACAGTGGTC TGTCAGCTAA AGTCGGAACG GGATGGAAG TGGTAACCTCT GTAGTGCAGA GGGGGTTCTA CGACAGAGGG

+2 R T Q R R G R T G R G K P G I Y R F V A P G E R P S G
 2721 GCACTCAACG TCGGGGAGG ACTGGCAGGG GGAAGGCCAGG CATCTACAGA TTGTGGCAC CGGGGGAGGG CCCCTCCGGC
 CGTAGTTGC AGCCCCGTCC TGACCGTCCC CCTCGGTCC GTAGATGTCT AAACACCGTG GCCCCCTCGC GGGGAGGGCC

+2 M F D S S V L C E C Y D A G C A W Y E L T P A E T T V
 2801 ATGTCGACT CGTCCGTCT CTGTGAGTGC TATGACGAG GCTGTGCTG GTATGAGCTC ACCCCCCGCG AGACTACAGT
 TACAAGCTGA GCAGGCAGGA GACACTCACG ATACTGCGTC CGACACGAAC CATACTCGAG TGCGGGCGGC TCTGATGTCA

+2 R L R A Y M N T P G L P V C Q D H L E F W E G V F T
 StuI
 2881 TAGGCTACGA CGGTACATGA ACACCCCGGG GCTTCCCGTG TGCCAGGACC ATCTGAATT TTGGGAGGGC GTCTTACAG
 ATCCGATGCT CGCATGTACT TGTGGGGCCC CGAAGGGCAC ACGGTCTGG TAGAACTTAA AACCTCCCG CAGAAATGTC

+2 G L T H I D A H F L S Q T K Q S G E N L P Y L V A Y Q
 StuI
 2961 GCCTCACTCA TATAGATGCC CACTTTCTAT CCCAGACAAA GCAGAGTGGG GAGAACCTTC CTACCTGGT AGCGTACCAA
 CGGAGTGAGT ATATCTACGG GTGAAAGATA GGGTCTGTT CGTCTCACCC CTCTGGAAG GAATGGACCA TCGCATGGTT

+2 A T V C A R A Q A P P P S W D Q M W K C L I R L K P T
 3041 GCCACCGTGT GCGCTAGGGC TCAAGCCCT CCCCATCGT GGGACAGAT GTGGAAAGTGT TTGATTGCC TCAAGCCAC
 CGGTGGCACA CGGATCCCG AGTCGGGGA GGGGTAGCA CCCTGTCTA CACCTTCACA AACTAAGGG AGTCGGGTG

+2 L H G P T P L L Y R L G A V Q N E I T L T H P V T K
 3121 CCTCCATGGG CCAACACCCC TGCTATACAG ACTGGGCGCT GTTCAGAATG AAATCACCC GACGCACCCA GTACCAAAAT
 GGAGGTACCC GGGTGTGGGG ACGATATGTC TGACCCGGCA CAAGTCTTAC TTTAGTGGGA CTGGTGTGGT CAGTGGTTA

+2 Y I M T C M S A D L E V V T S T W V L V G G V L A A L
 3201 ACATCATGAC ATGCATGTGCG GCGGACCTGG AGGTGTCAC GAGCACCTGG GTGTCGTTG GCGGCGTCT GGCTGCTTS
 TGTAGTACTG TACGTACAGC CGGCTGGACC TCCAGCAGTG CTCGTGGACC CACGAGCAAC CGCGCGCAGGA CGGACGAAAC

+2 A A Y C L S T G C V V I V G R V V L S G K P A I I P D
 3281 GCCGGTATT GCCTGTCAAC AGGCTGCGTG GTCATAGTGG GCAGGGTCGT CTTGTCGGG AAGCCGGCAA TCATACTGA
 CGGGCGATAA CGGACAGTT TCCGACGAC CAGTATCAC CGTCCCGAGCA GAACAGGCC TTCGGCCGT AGTATGGACT

+2 R E V L Y R E F D E M E E C S Q H L P Y I E Q G M M
 3361 CAGGGAAGTC CTCTACCGAG AGTCGATGA GATGGAAGAG TGCTCTCACG ACTTACCGTA CATCGAGCAA GGGATGATGC
 GTCCCTTCAG GAGATGGCTC TCAAGCTACT CTACCTTCTC ACGAGAGTC TGAAATGGCAT GTAGTCGTT CCCTACTACG

+2 L A E Q F K Q K A L G L L Q T A S R Q A E V I A P A V
 3441 TCGCCGAGCA GTCAAGCGAG AAGGCCCTCG GCCTCTGCA GACCGCGTCC CGTCAGGCAG AGTTATCGC CCCTGCTGTC
 ACCGGCTCGT CAAGTTCGTC TTCCGGAGC CGGAGGACGT CTGGCCGAGC GCAGTCGGTC TCCAATAGCG GGGACGACAG

+2 O T N W Q K L E T F W A K H M W N F I S G I Q Y L A G
 3521 CAGACCAACT GGCAAAACT CGAGACCTTC TGGCGAAGC ATATGTGGAA CTTCATCAGT GGGATACAAT ACTTGGCGGG
 GTCTGGTTGA CGTCTGGAA ACCCGCTCG TATACACCTT GAAGTAGTCA CCCTATGTTA TGAACCGCCCC

+2 L S T L P G N P A I A S L M A F T A A V T S P L T T
 3601 CTTGTCACG CTGGCTGGTA ACCCCGCAT TGCTTCATG ATGGCTTTA CAGCTGCTGT CACCAAGCCCA CTAACCACTA
 GAACAGTTGC GACGGACCAT TGGGGCGGTA ACGARAGTAAC TACCGAAAAT GTGACGACA GTGGTCGGGT GATTGGTGT

+2 S Q T L L F N I L G G W V A A Q L A A P G A A T A F V
 3681 GCCAAACCT CCTCTTCAC ATATTGGGGG GGTGGGGC TGCCCGCTC GCGGCCCCCG GTGCGCTAC TGCCTTGTG
 CGGTTGGGA GGAGAAGTTG TATAACCCCC CCACCCACCG ACGGGTCAG CGGGGGGGC CACGGCGATG ACGGAAACAC

pCMV-NS35

FIGURE 3 - Page 4

+2 G A G L A G A A I G S V G L G K V L I D I L A G Y G A
 3761 GGCCTGGCT TAGCTGGCG CGCCATCGGC AGTGTGGAC TGGGAAAGT CCTCATAGAC ATCCTGCAG GGTATGGCGC
 CGCGACCGA ATCGACCCCG GCGTAGCCG TCACAACCTG ACCCTTCCA GGAGTATCTG TAGAACGTC CCATACCGCG

+2 G V A G A L V A F K I M S G E V P S T E D L V N L L
 3841 GGGCTGGCG GGAGCTCTTG TGGCATTCAA GATCATGAGC GGTGAGGTCC CCTCCACGGA GGACCTGGTC AATCTACTGC
 CCCGCACCGC CCTCGAGAAC ACCGTAAGTT CTAGTACTCG CCACTCCAGG GGAGGTGCCT CCTGGACCAG TTAGATGACG

+2 P A I L S P G A L V V G V V C A A I L R R H V G P G E
 3921 CCCCATCCT CTCGCCCGA GCCCTCGTAG TCGCGTGGT CTGTGAGCA ATACTGCGCC GGACGTTGG CCCGGCGAG
 GCGTAGGA GAGCGGGCT CGGGAGCATC AGCCGACCA GACACGTCGT TATGACGCGG CGGTGCAACC GGGCCCGCTC

+2 G A V Q W M N R L I A F A S R G N H V S P T H Y V P E
 4001 GGGCAGTC AGTGGATGAA CGGCTGATA GCCTTCGCT CCCGGGGAA CCATGTTCC CCCACGCACT ACGTGCCGA
 CCCGTCACTG TCACCTACTT GGCGACTAT CGGAACGGA GGGCCCCCTT GGTACAAAGG GGGTGCSTGA TGACGGCCT

+2 S D A A A R V T A I L S S L T V T Q L L R R L H Q W
 4081 GAGCGATGCA GCTGCCCGC TCAC TGCCAT ACTCAGCAGC CTCAC GTAA CCCAGCTCT GAGGCGACTG CACCA GTGGCGA
 CTCGCTACGT CGACGGGCGC AGT GACGTTA TGAGTGTG GAGT GACATT GGGT GAGGA CTCCGCTGAC GTGGT CACCT

+2 I S S E C T T P C S G S W L R D I W D W I C E V L S D
 4161 TAAGCTGGA GTG TACCA CT CCATGCTCCG GTTCTGGCT AAGGGACATC TGGGACTGGA TATGAGGTT GTT GAGCGAC
 ATTCA GGCCT CACATGGTGA CGTACGAGGC CAAGGACCGA TTCCCTGTAG ACCCTGACCT ATACGCTCCA CAACTCGCTG

+2 F K T W L K A K L M P Q L P G I P F V S C Q R G Y K G
 BamHI

4241 TTTAAGACCT GGCTAAAAGC TAAGCTCATG CCACAGCTGC CTGGGATCCC CTTTGTGTCC TGCCAGCGCG GGTATAAGGG
 AAATTCTGGA CCGATTTCG ATTCGAGTAC GGTGTCGAGC GACCCTAGGG GAAACACAGG ACGGTGCAGC CCATATTCCC

+2 V W R G D G I M H T R C H C G A E I T G H V K N G T
 4321 GGTCTGGCA GGGGACGGCA TCATGCACAC TCGCTGCCAC TGTGGAGCTG AGATCACTGG ACATGTCAAA AACGGGAC
 CCAGACCGCT CCCCTGCCGT AGTACGTGTG AGC GACGGTG ACACCTCGAC TCTAGT GACC TTGCTCTGCT

+2 M R I V G P R T C R N M W S G T F P I N A Y T T G P C
 4401 TGAGGATCGT CGGCTCTAGG ACCTG CAGGA ACATGTGGAG TGGGACCTTC CCCATTAATG CCTACACCA C GGGCCCTGT
 ACT CCTAGCA GCCAGGATCC TGGACGTCT TGTACACCTC ACCCTGGAA GGTAATTAC GGATGTGGTG CCCGGGGACA

+2 T P L P A P N Y T F A L W R V S A E E Y V E I R Q V G
 4481 ACCCCCCCTC CTGCGCCGAA CTACAGTTG CCGCTATGGA GGGTGTCTG AGAGGAATAC GTGGAGATAA GGCAGGTGGG
 TGGGGGAG GACGCGGCTT GATGTGCAAG CGCGATACCT CCCACAGACG TCTCCTATG CACCTCTATT CGTCCACCC

+2 D F H Y V T G M T T D N L K C P C Q V P S P E F F T
 4561 GGACTTCCAC TACGTGACGG GTATGACTAC TGACAATCTT AAATGCCGT GCGAGGTCCC ATCGCCCGAA TTTTCACAG
 CCTGAAGGTG ATGCACTGCC CATACTGATG ACTGTTAGAA TTACGGGCA CGGTCAGGG TAGCGGGCTT AAAAAGTGT

+2 E L D G V R L H R F A P P C K P L L R E E V S F R V G
 4641 AATTGGACGG GGTGCGCTA CATAGTTTG CGCCCCCTG CAAGCCCTG CTGCGGGAGG AGGTATCATT CAGAGTAGGA
 TTACCTGCC CCACGCGGAT GTATCAAAC GCGGGGGAC GTTCGGGAAC GACCCCTCC TCCATAGTAA GTCTCATCT

+2 L H E Y P V G S Q L P C E P E P D V A V L T S M L T D
 4721 CTCCACGAAT ACCCGTAGG GTCGCAATT A CCTTGCGAGC CCGAACCGGA CGTGGCCGTG TTGACGTCCA TGCTCACTGA
 GAGGTGTTA TGCCCATCC CAGCGTTAAT GGAACGCTCG GGCTGGCCT GCACCGGCAC AACTGCAGGT ACGAGTGACT

+2 P S H I T A E A A G R R L A R G S P P S V A S S S A
 4801 TCCCTCCCAT ATAACAGCAG AGGCAGCGG GCGAAGGTG GCGAGGGAT CACCCCTC TGTGGCCAGC TCCTCGGCTA
 AGGGAGGGTA TATTGTCGTC TCCGCCGGCC CGCTCCCTA GTGGGGGGAG ACACCGGTG AGGAGCCGAT

pCMV-NS35

FIGURE 3 - Page 5

+2 S Q L S A P S L K A T C T A N H D S P D A E L I E A N
 4881 GCCAGCTATC CGCTCCATCT CTCAAGGCAA CTTGCACCGC TAACCATGAC TCCCCTGATG CTGAGCTCAT AGAGGCCAAC
 CGTCGATAG GCGAGGTAGA GAGTTCCGT GAACGTGGCG ATTGGTACTG AGGGGACTAC GACTCGAGTA TCTCCGGTTG

+2 L L W R Q E M G G N I T R V E S E N K V V I L D S F D
 4961 CTCTATGGA GGCAGGAGAT GGGCGCAAC ATCACCAAGG TTGAGTCAGA AAACAAAGTG GTGATTCTGG ACTCCTICGA
 GAGGATACT CCGCTCTA CCCGCCGTG TAGTGGTCC AACTCAGTCT TTTGTTTAC CACTAAGACC TGAGGAAGCT

+2 P L V A E E D E R E I S V P A E I L R K S R R F A Q
 5041 TCCGCTGTG GCGGAGGAGG ACGAGCGGGA GATCTCCGT CCCAGAAA TCCTGCGGAA GTCTCGGAGA TTGCGGCCAGG
 AGGCACAC CGCTCCTCC TGCTCGCCCT CTAGAGGCAT GGGCGCTTT AGGACGCCCT CAGAGCCTCT AACGGGGTCC

+2 A L P V . W A R P D Y N P P L V E T W . K K P D Y E P P V
 5121 CCTGCCCCGT TTGGGCGCGG CCGGACTATA ACCCCCCGCT AGTGGAGACG TGAAAAAACC CCGACTACGA ACCACCTGTG
 GGGACGGGCA AACCGCGC GGCTGATAT TGGGGGCGA TCACCTCTGC ACCTTTTTCG GGCTGATGCT TGGGGACAC

+2 V H G C P L P P P K S P P V P P P R K K R T V V L T E
 5201 GTCCATGGCT GCCCGCTTC ACCTCCAAAG TCCCTCTG TGCTCCGCC TCAGAAGAAG CGGACGGTGG TCCTCACTGA
 CAGGTACCGA CGGGCGAAGG TGGAGGTTTC AGGGGAGGAC ACGGAGGCGG AGCCTTCTTC GCCTGCCACC AGGAGTGA

+2 S T L S T A L A E L A T R S F G S S S T S G I T G D
 5281 ATCAACCTA TCTACTGCCT TGGCCGAGCT CGCCACCAAGA AGCTTGGCA GCTCTCAAC TTCCGGCATT ACGGGCGACA
 TAGTTGGAT AGATGACCGA ACCGGCTCGA GCGGTGGTCT TCGAAACCGT CGAGGAGTTC AAGGCCCTAA TGCCGCTGT

+2 N T T T S S E P A P S G C P P D S D A E S Y S S M P P
 5361 ATACGACAAC ATCCTCTGAG CCCGCCCCCT CTGGCTGCC CCCCCACTCC GACGCTGAGT CCTATTCTC CATGCCCTCC
 TATGCTGTTG TAGGAGACTC GGGCGGGAA GACCGACGGG GGGCTGAGG CTGCGACTCA GGATAAGGAG GTACGGGGGG

+2 L E G E P G D P D L S D G S W S T V S S E A N A E D V
 BamHI

5441 CTGGAGGGGG AGCTGGGAA TCCGGATCT AGCGACGGGT CATGGTCAAC GGTCAAGTAGT GAGGCCAACG CGGAGGATGT
 GACCTCCCCC TCGACCCCT AGGCCTAGAA TCGCTCCA GTACCAAGTTG CCAGTCATCA CTCCGGTGC GCCTCCTACA

+2 V C C S M S Y S W T G A L V T P C A A E E Q K L P I
 5521 CGTGTGCTGC TCAATGTCTT ACTCTTGAC AGGCGACTC GTCAACCGT GCGCCGCCGA AGAACAGAAA CTGCCCATCA
 GCACACGAGC AGTTACAGAA TGAGAACCTG TCCGGTGAG CAGTGGGGCA CGCGGCCCT TCTTGCTTT GACGGGTAGT

+2 N A L S N S L L R H H N L V Y S T T S R S A C Q R Q K
 5601 ATGCACTAAG CAACTCGTT CTACGTCAAC ACAATTGGT GTATTCACC ACCTCACCGA GTGCTGCCA AAGGCAGAAG
 TACGTGATTC GTTGGACAAAC GATGCAAGTGG TGTAAACCA CATAAGGTGG TGGAGTCCGT CACGAACGGT TTCCGTCTC

+2 K V T F D R L Q V L D S H Y Q D V L K E V K A A A S K
 5681 AAAGTCACAT TTGACAGACT GCAAGTTCTG GACACCCATT ACCAGGACGT ACTCAAGGAG GTAAAGCAG CGCGCTCAA
 TTTCAGTGTAA AACTGTCTGA CGTTCAAGAC CTGTCGGTAA TGGCTCTGCA TGAGTTCTC CAATTCTGC CGCGCAGTTT

+2 V K A N L L S V E E A C S L T P P H S A K S K F G Y
 5761 AGTGAAGGCT AACTTGCTAT CCGTAGAGGA AGCTTGACGC CTGACGCCCG CACACTCAGC CAAATCCAAG TTTGGTTATG
 TCACTCCGA TTGAACGATA GGCATCTCT TCGAACGTG GACTGGGGGG GTGTGACTCG GTTGTAGTTC AAACCAATAC

+2 G A K D V R C H A R K A V T H I N S V W K D L L E D N
 5841 GGGCAAAAGA CGTCCGTTGC CATGCCAGAA AGGCCTAAC CCACATCAAC TCCGTGTGGA AAGACCTCT GGAAGACAAT
 CCCGTTCTC CGAGGCAACG GTACGGCTT TCCGGCATTG GGTGTAGTTG AGGCACACCT TTCTGGAAGA CCTCTCTGTA

+2 V T P I D T T I M A K N E V F C V Q P E K G G R K P A
 5921 GTAACACCAA TAGACACTAC CATCATGGCT AAGAACGAGG TTTCTGCGT TCAGCCTGAG AAGGGGGTC GTAAGCCAGC
 CATTGTGGTT ATCTGTGATG GTAGTACCGA TTCTGCTCC AAAAGACGCA AGTCGGACTC TTCCCCCAG CATTGGTCC

pCMV-NS35

FIGURE 3 - Page 6

+2 R L I V F P D L G V R V C E K M A L Y D V V T K L P
 6001 TCGTCTCATC GTGTTCCCG ATCTGGCGT GCGCGTGCG AAAAGATGG CTTGTACGA CGTGGTTACA AGACTCCCCT
 AGCAGAGTAG CACAAGGGGC TAGACCCCA CGCGCACACG CTTTCTACC GAAACATGCT GCACCAATGT TTGAGGGGA

+2 L A V M 'G S S Y G F Q Y S P G Q R V E F L V Q A W K S
 EcoRI

6081 TGGCCGTGAT GGGAAAGCTCC TACGGATTCC AATACTCACC AGGACAGCGG GTTGAATTCC TCGTGCAGC GTGGAAGTCC
 ACCGGCACTA CCCTTCGAGG ATGCCCTAAGG TTATGAGTGG TCCTGTCGCC CAACTTAAGG AGCACGTTCG CACCTTCAGG

+2 K K T P M G F S Y D T R C F D S T V T E S D I R T E E
 6161 AAGAAAACCC CAATGGGGTT CTCGTATGAT ACCCGCTGCT TTGACTCCAC AGTCACTGAG AGCGACATCC GTACGGAGGA
 TTCTTTGGG GTTACCCCAA GAGCATACTA TGGCGACGA AACTGAGGTG TCAGTGAAC TCGTGTAGG CATGCCCTCCT

+2 A I Y Q C C D L D P Q A R V A I K S L T E R L Y V G
 6241 GGCAATCTAC CAATGTTGTG ACCTCGACCC CCAAGCCCGC GTGGCCATCA AGTCCCTCAC CGAGAGGCTT TATGTTGGGG
 CCGTTAGATG GTTACAACAC TGGAGCTGGG GGTTCCGGCG CACCGTAGT TCAGGGAGTG GCTCTCCGAA ATACAACCCC

+2 G P L T N S R G E N C G Y R R C R A S G V L T T S C G
 6321 GCCCTCTTAC CAATTCAGG GGGGAGAACT GCGGCTATCG CAGGTGCCGC GCGAGCGGCG TACTGACAAC TAGCTGTGGT
 CGGGAGAATG GTTAAGTTCC CCCCTTTGA CGCGATAGC GTCCACGGCG CGCTCGCCGC ATGACTGTTG ATGACACACCA

+2 N T L T C Y I K A R A A C R A A G L Q D C T M L V C G
 6401 AACACCCCTCA CTGCTCATAT CAAGGCCCCG GCAGCCTGTC GAGGCCAGG GCTCCAGGAC TGACCATGC TCGTGTGTGG
 TTGTGGAGT GAACGATGTA GTTCCGGGCC CGTCGGACAG CTCGGCTCC CGAGGTCTG ACGGTGGTACG AGCACACACC

+2 D D L V V I C E S A G V Q E D A A S L R A F T E A M
 6481 CGACGACTTA GTCTGTATCT GTGAAAGCGC GGGGGTCCAG GAGGACGGG CGAGCCTGAG AGCCTTCACG GAGGCTATGA
 GCTGCTGAAT CAGCAATAGA CACTTCGCG CCCCCAGGTC CTCCCTGCC GCTCGGACTC TCGGAAGTGC CTCCGATACT

+2 T R Y S A P P G D P P Q P E Y D L E L I T S C S S N V
 6561 CCAGGTACTC CGCCCCCCCCT GGGGACCCCC CACAACCAGA ATACGACTTG GAGTCATAA CATCATGTC CTCCAACGTG
 GGTCATGAG GCGGGGGGA CCCCTGGGG GTGTTGGTCT TATGCTGAAC CTGGAGTATT GTAGTACCGAG GAGGTTGCAC

+2 S V A H D G A G K R V Y Y L T R D P T T P L A R A A W
 6641 TCAGTCGCC ACCACGGCGC TGGAAAGAGG GTCTACTACC TCACCCGTGA CCCTACAACC CCCCTCGGA GAGCTGCCTG
 AGTCAGCGGG TGCTGCCGC ACCTTCTCC CAGATGATGG AGTGGCACT GGGATGTTGG GGGGAGCGCT CTCGACGCAC

+2 E T A R H T P V N S W L G N I I M F A P T L W A R M
 6721 GGAGACAGCA AGACACACTC CAGTCATTIC CTGGCTAGGC AACATAATCA TGTTGCCCT CACACTGTGG GCGAGGATGA
 CCTCTGTCGT TCTGTGTAG GTCAGTTAG GACCGATCCG TTGTATTAGT ACAAAAGGGG GTGTGACACC CGCTCCTACT

+2 I L M T H F F S V L I A R D Q L E Q A L D C E I Y G A
 6801 TACTGATGAC CCATTCTTT AGCGTCTTA TAGCCAGGG CCAGCTGAA CAGGCCCTCG ATTGCAGAT CTACGGGGCC
 ATGACTACTG GGTAAGAAA TCGCAGGAAT ATCGGTCCCT GGTCGAACCT GTCCGGGAGC TAACGCTCTA GATGCCCCGG

+2 C Y S I E P L D L P P I I O R L H G L S A F S L H S Y
 6881 TGCTACTCCA TAGAACACT GGATCTACCT CCAATCATTC AAAGACTCCA TGGCCTCAGC GCATTTCAC TCCACAGTA
 ACGATGAGGT ATCTGGTGA CCTAGATGGA GTTAGTAAG TTTCTGAGGT ACCGGACTCG CGTAAAAGTG AGGTGTCAAT

+2 S P G E I N R V A A C L R K L G V P P L R A W R H R
 6961 CTCTCCAGGT GAAATCAATA GGGTGGCCGC ATGCTCAGA AACTTGGGG TACCGCCCTT GCGAGCTTGG AGACACCGGG
 GAGAGGTCCA CTTAGTTAT CCCACCGGG TACGGAGTCT TTGAAACCCC ATGGCGGGAA CGCTCGAACCC TCTGTGGCCC

+2 A R S V R A R L L A R G G R A A I C G K Y L F N W A V
 7041 CCCGGAGCGT CGCGCCTAGG CTCTGGCCA GAGGGAGCAG GCGTGCCTATA TGTTGGCAAGT ACCTCTCAA CTGGGCAGTA
 GGGCCTOGCA GGCGCGATCC GAAGACCGGT CTCCCTCGTC CCGACGGTAT ACACCGTTCA TGGAGAAGTT GACCCGTAT

pCMV-NS35

FIGURE 3 - Page 7

+2 R T K L K L T P I A A A G Q L D L S G W F T A G Y S G
 7121 AGAACAAAGC TCAAACATCAC TCCAATAGCG GCCGCTGCC AGCTGGACTT GTCCGGCTGG TTCACGGCTG GCTACAGCGG
 TCTTGTTCG AGTTGAGTG AGGTTATCGC CGGGCACCGG TCGACCTGAA CAGGCCGACC AAGTGCCGAC CGATGTCGCC

+2 G D I Y H S V S H A R P R W I W F C L L L L A A G V
 7201 GGGAGACATT TATCACAGCG TGCTCATCC CGGGCCCCCG TGGATCTGGT TTGCTACT CTCGCTGCT GCAGGGTAG
 CCCCTGTAA ATAGTGTGCG ACAGAGTACG GGCGGGGCG ACCTAGACCA AACCGGATGA GGACGAACGA CGTCCCCATC

+2 G I Y L L P N R
 7281 GCATCTACCT CCTCCCCAAC CGATGAAGGT TGGGTAAC ACTCCGGCTT AAAAAAAA AAAATCTAG AAAGGCCGC
 CGTAGATGGA GGAGGGGTTG GCTACTTCCA ACCCCATTG TGAGGCCGA TTTTTTTTT TTTTAGATC TTTCGGCGG

BamHI MluI

7361 CAAGATATCA AGGATCCACT ACGCGTTAGA GCTCGCTGAT CAGCCCTCGAC TGTGCCCTCT AGTTGCCAGC CATCTGTTGT
 GTTCTATAGT TCCTAGGTGA TGCGCAACT CGAGCGACTA GTCGGAGCTG ACACGGAAAGA TCAACGGTGG TAGACAACA

7441 TTGCCCCCTCC CCCGTGCCTT CCTTGACCTT GGAAGGTGCC ACTCCCACTG TCTTCCCTA ATAAAATGAG GAAATGGAT
 AACGGGGAGG GGGCACGGAA GGAACGGGA CCTTCCACCG TGAGGGTGAC AGGAAAGGAT TATTTACTC CTTAACGTA

7521 CGCATTGTCT GAGTAGGTGT CATTCTATTC TGGGGGGTGG GGTGGGGCAG GACAGCAAGG GGGAGGATTG GGAAGACAT
 CGCTAACAGA CTCATCCACA GTAAGATAAG ACCCCCCACC CCACCCCGT CTGTCGTTCC CCCCCTAAAC CTTCTGTAA

7601 AGCAGGCATG CTGGGGAGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTGCGT TCGGCTGCCG CGAGCGGTAT
 TCGTCCGTAC GACCCCTCGA GAAGGGAAAG GAGGGAGTGA CTGAGCGACG CGAGCGACGCC GCTCGCCATA

7681 CAGCTCACTC AAAGGCGGTAA ATACGGTTAT CCACAGAAC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG
 GTCGAGTGAG TTTCCGCAT TATGCCATA GGTGTCTTAG TCCCCTATTG CGTCCTTCT TGTACACTCG TTTCCGGTC

7761 CAAAGGCCA GGAACCGTAA AAAGGCCCGG TTGCTGGCGT TTTCCATAG GCTCCGGCCC CCTGACGAGC ATCACAAAA
 GTTTCCGGT CCTTGGCATT TTTCCGGCG AACGACCGCA AAAGGTATC CGAGGCCGGG GGACTGCTCG TAGTGTTTT

7841 TCGACGCTCA AGTCAGAGGT GGCAGAACCC GACAGGACTA TAAAGATACC AGGGCTTCC CCGTGAAGC TCCCTCGTGC
 AGCTGGAGT TCACTCTCCA CGCTTGGG CTGCTGTAT ATTCTATGG TCCGCAAAGG GGGACCTTCG AGGGAGCACG

7921 GCTCTCTGT TCCGACCTCG CCGCTTACCG GATACCTGTC CGCTTCTC CCTCAGGAA GCGTGGCGCT TTCTCAATGC
 CGAGAGGACA AGGCTGGAC GGCAGATGCC CTATGGACAG GCGGAAAGAG GGAAGCCCT CGCACCGCA AAGAGTTACG

8001 TCACTGCTGA GGATCTCAAG TTGGTGTAG GTCTGGCTC CCAAGCTGGG CTGTCGAC GAACCCCGG TTGAGCCGA
 AGTGCACAT CCATAGAGTC AAGCCACATC CAGCAAGCGA GTTCAACCC GACACACGT CTTGGGGC AAGTGGGCT

8081 CCGCTGCCCT TTATCCGGTA ACTATCGTCT TGAGTCAAC CGGTAAGAC AGCACTTATC GCCACTGGCA GCAGCCACTG
 GGCAGCCGG AATAGCCAT TGATAGCAGA ACTCAGGTG GGCAATTCTG TGCTGAATAG CGGTGACCGT CGTCGGTAC

8161 GTCACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACACTGG CTACACTAGA
 CATTGTCTTA ATCGTCTCGC TCCATACATC CGCCACGATG TCTCAAGAAC TTCACCCACG GATTGATGCC GATGTGATCT

8241 AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA CCTTCGAAA AAGAGTTGGT AGCTCTTGAT CGGGCAAAACA
 TCTCTGTATA AACCATAGAC GCGAGACGAC TTGGTCAAT GGAAGCCCTT TTCTCAACCA TCGAGAACTA GGCGTTGCT

8321 AACCAACCGCT GGTAGCGGTG GTTTTTGT TTGCAAGCG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT
 TTGGTGGCGA CCATGCCAC CAAAAAAACA AACGTCGTC GTCTAATGCG CGTCTTTTT TTCTAGAGTT CTTCTAGGAA

8401 TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAAA CTCACGTTAA GGGATTTGG TCATGAGATT ATCAAAAGG
 ACTAGAAAAG ATGCCAGA CTGGAGTC CTTGGCTTT GAGTGCATT CCCTAAAACC AGTACTCTAA TAGTTTCC

pCMV-NS3S

FIGURE 3 - Page 8

8481 ATCTTCACCT AGATCCTTT AAATTAAGGGAA TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG
TAGAAGTGGA TCTAGGAAAA TTAAATTGTT ACTTCAAAT TTAGTTAGT TTCATATATA CTCATTTGAA CCAGACTGTC

8561 TTACCAATGC TTATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTG GTTCATCCAT AGTTGCCTGA CTCCCCGTGCG
AATGGTTACG AATTAGTCAC TCCGTGGATA GAGTCGCTAG ACAGATAAG CAAGTAGGT ACAAACGGACT GAGGGGCAGC

8641 TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGCCTGCA ATGATACCGC GAGACCCACG CTCACCGGCT
ACATCTATTG ATGCTATGCC CTCCCGAATG GTAGACCGGG GTCACGACGT TACTATGGCG CTCTGGTGC GAGTGGCCGA

8721 CCAGATTAT CAGCAATAAA CCAGGCCAGCC GGAAGGGCCG AGCGCAGAAG TGTTCCCTGCA ACTTTATCCG CCTCCATCCA
GGTCTAAATA GTCGTTATTT GGTCGGTCCG CCTTCCCGGC TCGCTCTTC ACCAGGACGT TGAAATAGGC GGAGGTAGGT

8801 GTCTATTAAT TGTGCCCCGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCCTAA CGTTGTTGCC ATTGCTACAG
CAGATAATTA ACAACGGCC TTCGATCTCA TTCATCAGC GTCATTAT CAAACCGGT GCAACACGG TAACGATGTC

8881 GCATCGTGGT GTCACGCTCG TCGTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGGCAGT TACATGATCC
CGTAGCACCA CAGTGCAGGC AGCAAACCAT ACCGAGTAA GTCGAGGCCA AGGGTTGCTA GTTCCGCTA ATGTAAGTGG

8961 CCCATGTTGT GCAAAAAGC GGTTAGCTCC TTGGTCCCTC CGATCGTTGT CAGAAGTAAG TTGGCCGAG TGTTATCACT
GGGTACAACA CGTTTTTCG CCAATCGAGG AAGCCAGGAG GCTAGCAACA GTCTCATTC AACCGGCCTC ACAATAGTGA

9041 CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCCGTA GATGCTTTG TGTTGACTGGT GAGTACTCAA
GTACCAATAC CGTCTGACG TATTAAGAGA ATGACAGTAC GGTAGGCATT CTAGAAAAG AACTGACCA CTATCAGTT

9121 CCAAGTCATT CTGAGAAATAG TGTATGCCG GACCGAGTTG CTCTGCCCG GCGTCAATAC GGATAATAC CGGCCACAT
GGTCAGTAA GACTCTTATC ACATACGCCG CTGGCTCAAC GAGAACGGGC CGCAGTTATG CCTTATTATG GCGCGGTGTA

9201 AGCAGAACTT TAAAAGTGT CATATTGGA AACCGTTCTT CGGGGGAAA ACTCTCAAGG ATCTTACCCG TGTTGAGATC
TCGTTTGAA ATTTTACAGA GTAGTAACTT TTGCAAGAA GCCCCGTT TGAGAGTCC TAGAATGGCG ACAACTCTAG

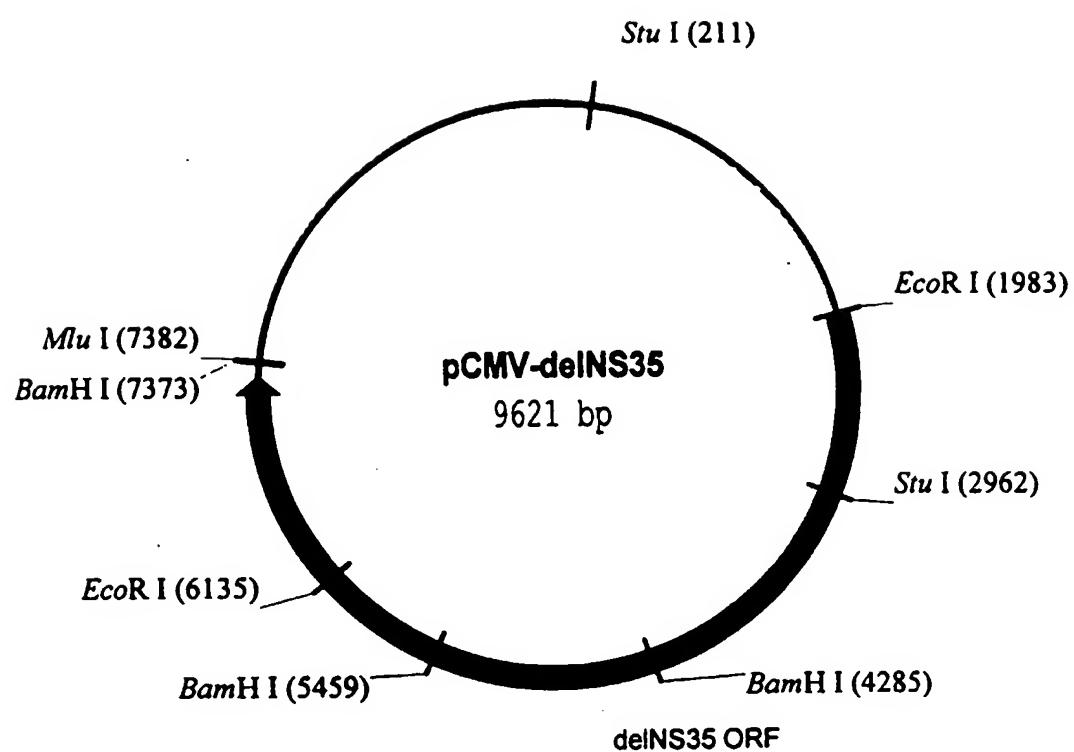
9281 CAGTCGATG TAACCCACTC GTGCACCCAA CTGATCTTC GAATCTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAA
GTCAGCTAC ATTGGGTAG CAGTGGGT GACTAGAAGT CGTAGAAAAT GAAAGTGGTC GCAAAGACCC ACTCGTTTT

9361 CAGGAAGGCA AAATGCCGCA AAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCTT TTTCAATAT
GTCCTTCCGT TTACGGCGT TTTTCCCTT ATTCCCGCTG TGCCTTACA ACTTATGAGT ATGAGAAGGA AAAAGTTATA

9441 TATTGAAGCA TTATCAGGG TTATGCTCTC ATGAGCGGT ACATATTGA ATGTATTTAG AAAATAAAC AAATAGGGT
ATAACTTCGT AAATAGTCCC AATAACAGAG TACTCGCTA TGATAAATC TACATAAATC TTTTATTTG TTTATCCCCA

9521 TCCGGCACA TTCCCCGAA AAGTGCCACC TGACGTCTAA GAAACCATTA TTATCATGAC ATTAACCTAT AAAAATAGGC
AGGCGCGTGT AAAGGGGCTT TTCACGGTGG ACTGCAGATT CTTTGGTAAT AATAGTACTG TAATTGGATA TTTTATCCG

9601 GTATCACGAG GCCCTTTCGT C
CATAGTGCCTC CGGGAAAGCA G

FIGURE 4

pCMV-de1NS35

FIGURE 5 - Page 1

1 TCGCGCGTTT CGGTGATGAC GGTAAAACC TCTGACACAT GCAGCTCCG GAGACGGTCA CAGCTTGTCT GTAAAGCGGAT
AGCGCGAAA GCCACTACTG CCACCTTG AGACTGTGTA CGTCGAGGGC CTCTGCCAGT GTCGAACAGA CATTGCGCTA

81 GCGGGGAGCA GACAAGCCCG TCAGGGCGC TCAGCGGGTG TTGGGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA
CGGCCCTCGT CTGTTGGGC AGTCCCGCG AGTCGCCAC AACCGCCAC AGCCCCGACC GAATTGATAC CCCGTAGTCT

StuI

161 GCAGATTGTA CTGAGAGTGC ACCATATGAA GCTTTTGCA AAAGCCTAGG CCTCCAAAAA AGCCTCCTCA CTACTTCTGG
CGTCTAACAT GACTCTCAGG TGGTATACTT CGAAAAACGT TTTCGGATCC GGAGGTTTT TCGGAGGAGT GATGAAGACC

241 AATAGCTCAG AGGCCGAGGC GGCCCTCGGC TCTGCATAAA TAAAAAAAAT TAGTCAGCCA TGGGGCGGAG AATGGGGAG
TTATCGAGTC TCCGGCTCCG CCGGAGCCGG AGACGTATT ATTTCCTTA ATAGTCGGT ACCCCGCCTC TTACCCGCCT

321 ACTGGGCGGG GAGGGAAATTA TTGGCTATTG GCCATTGCA ACCTGTTATC TATATCATAA TATGTACATT TATATTGGCT
TGACCCGCC CTCCTTAAT AACCGATAAC CGGTAACGTA TGCAACATAG ATATAGTATT ATACATGTAATATAACCGA

401 CATGCTCAAT ATGACCGCCA TGTGACATT GATTATTGAC TAGTTATAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT
GTACAGGTTA TACTGGCGT ACAACTGTAA CTAAATACTG ATCAATAATT ATCATTAGTT ATGCCCCAG TAATCAAGTA

481 AGCCCATAA TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCGCC TGGCTGACCG CCCAACGACC CCCGCCATT
TCGGGTATAT ACCTCAAGGC GCAATGTATT GAATGCCATT TACCGGGCGG ACCGACTGGC GGTTGCTGG GGGGGGTTAA

561 GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCATAA GGGACTTCC ATTGACGTCA ATGGTGGAG TATTTACGGT
CTGCAGTTAT TACTGCATAC AAGGGTATCA TTGCGGTTAT CCCTGAAAGG TAACTGCAGT TACCCACCTC ATAATGCCA

641 AAACGCCCA CTGGCAGTA CATCAAGTGT ATCATATGCC AAGTCCGCC CCTATTGACG TCAATGACGG TAAATGGGCC
TTGACGGGT GAAACCGTCA GTAGTTCACA TAGTATACGG TTCAGGGGG GGATAACTGC AGTTACTGCC ATTACCGGG

721 GCCTGGCATT ATGCCAGTA CATGACCTTA CGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC
CGGACCGTAA TACGGGTCA GTACTGGAAT GCCCTGAAAG GATGAACCGT CATGTAGATG CATAATCAGT AGCGATAATG

801 CATGGTATG CGGTTTGGC AGTACACCAA TGGCGTGG A TAGCGTTTG ACTCACGGG ATTTCAGT CTCCACCCCA
GTACCACTAC GCAAAACCG TCATGTGGTT ACCCGCACCT ATGCCAAAC TGAGTGGCCC TAAAGGTTCA GAGGTGGGGT

881 TTGACGTCAA TGGGAGTTG TTTGGCACC AAAATCAACG GGACTTCCA AAATGTGTA ATAACCCCGC CCCGTTGACG
AACTGCAGTT ACCCTCAAAAC AAAACCGTGG TTTAGTGTG CCTGAAAGGT TTACAGCAT TATTGGGGCG GGGCAACTGC

961 CAAATGGGCG GTAGGCGTGT ACGGTGGAG GTCTATATAA GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG
GTTACCCGC CATCCGACA TGCCACCTC CAGTATATT CGTCTCGAGC AAATCACTG GCAGTCTAGC GGACCTCTGC

1041 CCATCCACCG TGTTTGACCC TCCATAGAAG ACACCGGAC CGATCCAGCC TCCGGCCCG GGAACGGTGC ATTGGAACGC
GGTAGGTGCG ACAAAACTGG AGGTATCTTC TGTGCCCTG GCTAGGTGCG AGGCACCGC CCTTGCACAG TAACCTTGCG

1121 GGATTCCCCG TGCCAAGAGT GACGTAAGTA CGGCCATAG ACTCTATAGG CACACCCCTT TGGCTTTAT GCATGCTATA
CCTAAGGGC ACGGTTCTCA CTGCAATTCA GGCAGATATC TGAGATATCC GTGTGGGAA ACCGAGAATA CGTACGATAT

1201 CTGTTTTGG CTTGGGGCCT ATACACCCCG GCTCCTTATG CTATAGGTGA TGGTATAGCT TAGCCTATAG GTGTGGGTTA
GACAAARACC GAAACCCGGG TATGTGGGG CGAGGAATAC GATATCCACT ACCATATCGA ATCGGATATC CACACCCAAAT

1281 TTGACCATTA TTGACCACTC CCCTATTGGT GACGATACTT TCCATTAATCA ATCCATAACA TGGCTTTG CCACAACTAT
AACTGGTAAT AACTGGTGAG GGGATAACCA CTGCTATGAA AGGTATGAT TAGGTATTGT ACCGAGAAC GGTGTTGATA

1361 CTCTATTGGC TATATGCCAA TACTCTGTCC TTCAGAGACT GACACGGACT CTGTATTTT ACAGGATGGG GTCCATTAT
GAGATAACCG ATATACGGT ATGAGACAGG AAGTCTCTGA CTGTGCCCTGA GACATAAAA TGTCTACCC CAGGTAAATA

pCMV-delNS35

FIGURE 5 - Page 2

1441 TATTTACAAA TTCACATATA CAACAACGCC GTCCCCGTG CCCGAGTTT TTATTAACA TAGCGTGGGA TCTCCGACAT
ATAAATGTTT AAGTGTATAT GTTGTGCGG CAGGGGCAC GGGCGTCAA AAATATTGT ATCGCACCT AGAGGCTGTA

1521 CTCGGGTAAG TGTTCGGAC ATGGGCTCTT CTCCGGTAGC GGCGGAGCTT CCACATCCGA GCCCTGGTCC CATCCGTCCA
GAGCCCATGC ACAAGGCCTG TACCCGAGAA GAGGCCATCG CGGCCTCGAA GGTGTAGGCT CGGGACCAGG GTAGGCAGGT

1601 GCGGCTCATG GTCGCTCGGC AGCTCCTTG TCCTAACAGT GGAGGCCAGA CCTAGGCACA CCACAATGCC CACCACCA
CGCCGAGTAC CAGCGAGCCG TCGAGGAACG AGGATTGTCA CCTCCGGTCT GAATCCGTGT CGTGTACGG GTGGTGGTGG

1681 AGTGTGCGGC ACAAGGCCGT GGCGGTAGGG TATGTGTCTG AAAATGAGCT CGGAGATTGG GCTCGCACCT GGACGCAGAT
TCACACGGGC TGTTCCGGCA CGGCCATCCC ATACACAGAC TTTTACTCGA GCCTCTAACCG CGAGCGTGGA CCTGGTGTCA

1761 GGAAGACTTA AGGCAGCGGC AGAAGAAGAT GCAGGCAGCT GAGTGTGT ATTCTGATAA GAGTCAGAGG TAATCCCGT
CCTCTGAAT TCCGTGCGG TCTTCTCTA CGTCCGTGCA CTCAACRACA TAAGACTATT CTCAGTCTCC ATTGAGGGCA

1841 TGCGGTGCTG TAAACGGTGG AGGGCAGTGT AGTCTGAGCA GTACTCGTTG CTGCCGCCG CGCCACCAGA CATAATAGCT
ACGCCACGAC AATTGCCACC TCCCGTCACA TCAGACTCGT CATGAGCAAC GACGGCGCGC CGCGTGGTCT GTATTATCGA

+2 M A A
EcoRI

1921 GACAGACTAA CAGACTGTT CTTTCATGG GTCTTTCTG CAGTCACCGT CGTCACCTA AGAATTCAACC ATGGCTGCAT
CTGTCGATT GTCTGACAG GAAAGTACG CAGAAAAGAC GTCAGTGGCA GCAGTGGAT TCTTAAGTGG TACCGACGTA

2001 +2 Y A A Q G Y K V L V L N P S V A A A T L G F G A Y M S K
ATGCAGCTCA GGGCTATAAG GTGCTAGTAC TCAACCCCTC TGTTGCTGCA ACACTGGGCT TTGGTCTTA CATGTCACAG
TACGTCGAGT CCCGATATTC CACGATCATG AGTGGGGAG ACAACGACGT TGTGACCGA AACACGAAAT GTACAGGTT

2081 +2 A H G I D P N I R T G V R T I T T G S P I T Y S T Y G
GCTCATGGGA TCGATCCTAA CATCAGGACC GGGGTGAGAA CAATTACACAC TGCGAGCCCC ATCACGTACT CCACCTACGG
CGAGTACCCCT AGCTAGGATT GTAGTCCTGG CCCACTCTT GTTAATGGTG ACCGTCGGGG TAGTGCATGA GGTGGATGCC

2161 +2 K F L A D G G C S G G A Y D I I I C D E C H S T D A
CAAGTTCCCTT GCCGACGGCG GGTGCTCGGG GGGCGCTTAT GACATAATAA TTGTCGACGA GTGCCACTCC ACGGATGCC
GTTCAAGGAA CGGCTGCCGC CCACGAGCCC CCCGGAATA CTGTATTATT AAACACTGCT CACGGTGAGG TGCCCTACGGT

2241 +2 T S I L G I G T V L D Q A E T A G A R L V V L A T A T
CATCCATCTT GGGCATGGC ACTGTCCTTG ACCAAGCAGA GACTGCGGGG GCGAGACTGG TTGTGCTCGC CACCGCCACC
GTAGGTAGAA CCCGTAACCC TGACAGGAAC TGTTGCTCT CTGACGCCCG CGCTCTGACC AACACGAGCG GTGGCGGTGG

2321 +2 P P G S V T V P H P N I E E V A L S T T G E I P F Y G
CCTCCGGGCT CCGTCACTGT GCCCCATCCC AACATCGAGG AGGTGCTCT GTCCACCAAC GGAGAGATCC CTTTTACGG
GGAGGCCCGA GGCAGTGACA CGGGTAGGG TTGAGCTCC TCCACGAGA CAGGTGGTGG CCTCTCTAGG GAAAATGCC

2401 +2 K A I P L E V I K G G R H L I F C H S K K K C D E L
CAAGGCTATC CCCCTCGAAG TAATCAAGGG GGGGAGACAT CTCATCTCT GTCATTCAAA GAAGAAGTGC GACGAACCTG
GTTCCGATAG GGGGAGCTTC ATTAGTCCC CCCCTCTGTA GAGTGAAGA CAGTAAGTTT CTTCTCACG CTGCTTGAGC

2481 +2 A A K L V A L G I N A V A Y Y R G L D V S V I P T S G
CCGCAAAGCT GGTGCGATTG GGCATCAATG CCGTGGCTA CTACCGCGGT CTGACGTGT CGTCATCCC GACCAAGGGC
GGCCTTCGA CCAGCTAAC CCGTAGTTAC GGCACCGGAT GATGGCGCCA GAACTGCACA GGCAGTAGGG CTGGTGGCCG

2561 +2 D V V V V V A T D A L M T G Y T G D F D S V I D C N T C
GATGTTGTCG TCGTGGCAAC CGATGCCCTC ATGACCGCT ATACCGCGA CTTCGACTCG GTGATAGACT GCAATACGTG
CTACACAGC AGCACCGTG GCTACGGAG TACTGGCGA TATGGCCGCT GAAGCTGAGC CACTATCTGA CGTTATGCAC

pCMV-de1NS35

FIGURE 5 - Page 3

+2 V T Q T V D F S L D P T F T I E T I T L P Q D A V S
 2641 TGTCACCCAG ACAGTCGATT TCAGCCTGAA CCCTACCTTC ACCATGAGA CAATCAGGT CCCCCAAGAT GCTGTCTCCC
 ACAGTGGTC TGTCAGCTAA AGTCGGAACG GGGATGGAAG TGGTAACCT GTAGTGCGA GGGGTTCTA CGACAGAGGG

+2 R T Q R R G R T G R G K P G I Y R F V A P G E R P S G
 2721 GCACTCAACG TCGGGGCAGG ACTGGCAGGG GGAAGCCAGG CATCTACAGA TTTGTGGCAC CGGGGGAGCG CCCCTCCGGC
 CGTAGTTGC AGCCCCGTCC TGACCGTCCC CCTCGGTCC GTAGATGTCT AACACCCGTG GCCCCCTCGC GGGGAGGCCG

+2 M F D S S V L C E C Y D A G C A W Y E L T P A E T T V
 2801 ATGGTCGACT CGTCCTGCT CTGTGAGTGC TATGACCGAG GCTGTGCTTG GTATGAGCTC ACGCCCGCCG AGACTACAGT
 TACAAGCTGA GCAGGCAGGA GACACTCAGG ATACTCGCTC CGACACGAAC CATACTCGAG TGCGGGCGGC TCTGATGTCA

+2 R L R A Y M N T P G L P V C Q D H L E F W E G V F T
 StuI
 --

2881 TAGGCTACGA GCGTACATGA ACACCCCGGG GCTTCCCGTG TGCCAGGACC ATCTTGAATT TTGGGAGGGC GTCTTACAG
 ATCCGATGCT CGCATGACT TGTGGGCC CGAAGGGCAG ACGGTCTGG TAGAACCTAA AACCCCTCCCG CAGAAATGTC

+2 G L T H I D A H F L S Q T K Q S G E N L P Y L V A Y Q
 StuI
 --

2961 GCCCTACTCA TATAGATGCC CACTTCTAT CCCAGACAAA GCAGAGTGGG GAGAACCTTC CTTACCTGGT AGCGTACCAA
 CGGAGTGAGT ATATCTACGG GTGAAAGATA GGGTCTGTT CGTCTCACCC CTCTTGAAG GAATGGACCA TCGCATGGT

+2 A T V C A R A Q A P P P S W D Q M W K C L I R L K P T
 3041 GCCACCGTGT GCGCTAGGGC TCAAGCCCT CCCCATCGT GGGACCGAT GTGGAAGTGT TTGATTGCC TCAAGCCAC
 CGTGGCACA CGCGATCCCG AGTTCGGGA GGGGTAGCA CCCTGGCTA CACCTCACCA AACTAAGCGG AGTTGGGTG

+2 L R G P T P L L Y R L G A V Q N E I T L T H P V T K
 3121 CCTCCATGGG CCAACACCCC TGCTATACAG ACTGGGGC GTTCAGAAAT AAATCACCT GACGCACCCA GTCACCAAT
 GGAGGTACCC GTTGTGGGG ACATATGTC TGACCCGCA CAAGTCTAC TTTAGTGGGA CTGCGTGGGT CAGTGGTTA

+2 Y I M T C M S A D L E V V T S T W V L V G G V L A A L
 3201 ACATCATGAC ATGCATGTCG GCGCACCTGG AGGTCGTCAC GAGCACCTGG GTGCTCGTT CGGGCGTCCT GGCTGCTTG
 TGTAGTACTG TACGTACAGC CGGCTGGACC TCCAGCAGTG CTCGTGGACC CACGAGCAAC CGCCGCAAGA CCGACGAAAC

+2 A A Y C L S T G C V V I V G R V V L S G K P A I I P D
 3281 GCCCGTATT GCCTGTCAAC AGGCTGGTG GTCATAGTGG GCAGGGCTGT CTGTCCGGG AAGCCGGCAA TCATACCTGA
 CGCGCATAA CGGACAGTTG TCCGACGAC CAGTATCAC CGTCCAGCA GAACAGGCC TTGGCCGTT AGTATGGACT

+2 R E V L Y R E F D E M E E C S Q H L P Y I E Q G M M
 3361 CAGGGAAAGTC CTCTACCGAG AGTCGATGA GATGGAAAGAG TGCTCTCAGC ACTTACCGTA CATGAGCAA GGGATGATGC
 GTCCCTCAG GAGATGGCTC TCAAGCTACT CTACCTCTC ACGAGAGTCG TGAATGGCAT GTAGCTCGTT CCCTACTACG

+2 L A E Q F K Q K A L G L L Q T A S R Q A E V I A P A V
 3441 TCGCGAGCA GTTCAGCAG AAGGCCCTG GCCTCTGCA GACCGCGTCC CGTCAGGAG AGGTTATCGC CCCTGCTGTC
 AGCGGCTCGT CAAGTCGTC TTCCGGGAGC CGGAGGACGT CTGGCGCAGG GCAGTCCGTC TCCAATAGCG GGGACGACAG

+2 Q T N W Q K L E T F W A K H M W N F I S G I Q Y L A G
 3521 CAGACCAACT GCGAAAAACT CGAGACCTTC TGGGCAAGC ATATGTGGAA CTTCATCAGT GGGATACAAT ACTTGGCGGG
 GTCTGGTTGA CCGTTTTGA CCTCTGGAAG ACCCCCTCG TATACACCTT GAAGTAGTCR CCCTATGTTA TGAACCGCCC

+2 L S T L P G N P A I A S L M A F T A A V T S P L T T
 3601 CTGTCAACG CTGCTGGTA ACCCCGCAT TGCTCATG ATGGCTTTA CAGCTGCTGT CACCGCCCA CTAACCACTA
 GAACAGTGC GACGGACCAT TGGGGCGTA ACAGAAGAAC TACCGAAAAT GTGACGACA GTGGTCGGGT GATTGGTGAT

+2 S Q T L L F N I L G G W V A A Q L A A P G A A T A F V
 3681 GCGAAACCT CCTCTTCAAC ATATTGGGG GGTGGTGGC TGCCCAGCTC GCGGCCCGCG GTGCCGCTAC TGCCCTTGTC
 CGGTTGGGA GGAGAGTTG TATAACCCCC CCACCCACCG ACGGTGCAG CGCCGGGGC CACGGGATG ACGGAAACAC

pCMV-delNS35

FIGURE 5 - Page 4

+2 G A G L A G A A I G. S V G L G K V L I D I L A G Y G A
 3761. GGCCTGGCT TAGCTGGCG CGCCATCGGC AGTGTGGAC TGGGAAAGGT CCTCATAGAC ATCCTTGAG GGTATGGCG
 CGCGCACCGA ATCGACCGG CCGTAGCCG TCACAACCTG ACCCTTCCA GGACTATCTG TAGGAACGTC CCATACCGCG

+2 G V A G A L V A F K I M S G E V P S T E D L V N L L
 3841 GGCCTGGCG GGAGCTCTTG TGGCATTCAA GATCATGAGC CGTGAGGTCC CCTCCACCGA GGACCTGGTC AATCTACTGC
 CCCGACCCG CCTCGAGAAC ACCGTAAGIT CTAGTACTCG CCACTCCAGG GGAGGTGCCT CCTGGACCGAG TTAGATGAGC

+2 P A I L S P G A L V V G V V C A A I L R R H V G P G E
 3921 CGGCCATCCT CTCGCCCCGA GCCCTCGTAG TCGGCGTGGT CTGTGAGCA ATACTGCGCC GGACGTTGG CCCGGGGAG
 GGCCTGGAGGA GAGCGGGCT CCGGAGCATC AGCCGACCA GACACGTCGT TATGACGCGG CGGTGCAACC GGGCCCCCTC

+2 G A V Q W M N R L I A F A S R G N H V S P T H Y V P E
 4001 GGGCAGTGC AGTGGATGAA CCGGCTGATA GCCTTGCCT CCGGGGGAA CCTATGTTCC CCCACGCACT ACCTGCGGA
 CCCCGTCACG TCACCTACTT GGCGACTAT CGGAAGCGGA GGGCCCCCTT GTACAAAGG GGGTGCCTGA TGACCGCCT

+2 S D A A A R V T A I L S S L T V T Q L L R R L H Q W
 4081 GAGCGATGCA GCTGCCCGG TCACTGCCAT ACTCAGCAGC CTCACTGAA CCCAGCTCCT GAGGCAGTG CACCACTGGA
 CTCGCTACGT CGACGGCGC AGTGACGGA TGAGTCGTCG GAGTGACATT GGGTCGAGGA CTCCGCTGAC GTGGTCACCT

+2 I S S E C T T P C S G S W L R D I W D W I C E V L S D
 4161 TAAGCTCGGA GTGTACCACT CCATGCTCGG GTTCTGGCT AAGGGACATC TGGGACTGGA TATGCGAGGT GTTGAGGCAC
 ATTCGAGCT CACATGGTGA GGTACGAGGC CAAGGACCGA TTCCCTGTAG ACCCTGACCT ATACGCTCCA CAACTCGCTG

+2 F K T W L K A K L M P Q L P G I P F V S C Q R G Y K G
 BamHI

4241 TTAAGACCT GGCTAAAAGC TAAGCTCATG CCACAGCTGC CTGGGATCCC CTTTGTGTCC TGCCAGCGG GGTATAAGGG
 AAATTCTGGA CCGATTTCG ATTCGAGTAC GGTGTCGAG GACCCTAGGG GAAACACAGG ACGGTCGCC CCATATTCCC

+2 V W R G D G I M H T R C H C G A E I T G H V K N G T
 4321 GGTCTGGCGA GGGGACGGCA TCATGCACAC TCGCTGCCAC TGTTGGACTG AGATCACTGG ACATGTCAA AACGGGACGA
 CGAGACCGCT CCCCTGCCGT AGTACGTGTG AGCGACGGT ACACCTCGAC TCTAGTGCCT TGTACAGTT TTGCCCTGCT

+2 M R I V G P R T C R N M W S G T F P I N A Y T T G P C
 4401 TGAGGATCGT CGGTCTAGG ACCTGCAGGA ACATGTGGAG TGGGACCTTC CCTACACAC GGGCCCTGT
 ACTCTAGCA GCCAGGATCC TGGACGTCT TGTACACCTC ACCCTGGAAG GGGTAATTAC GGATGTGGT CCCGGGGACA

+2 T P L P A P N Y T F A L W R V S A E E Y V E I R Q V G
 4481 ACCCCCCCTTC CTGCGCCGAA CTACACGTTG GCGCTATGGA GGGTGTCTGC AGAGGAATAC GTGGAGATAA GGCAGTGGG
 TGGGGGAAG GACCGGGCTT GATGTGCAAG CGCGATAACCT CCCACAGACG TCTCCTTATG CACCTCTATT CCGTCCACCC

+2 D F H Y V T G M T T D N L K C P C Q V P S P E F F T
 4561 GGACTTCCAC TACGTGACGG GTATGACTAC TGACAATCTT AAATGCCGT GCCAGGTCCC ATCGCCCGAA TTTTCACAG
 CCTGAAGGTG ATGCACTGCC CATACTGATG ACTGTAGAA TTACGGGCA CGGTCCAGGG TAGCGGGCTT AAAAGTGTGTC

+2 E L D G V R L H R F A P P C K P L L R E E V S F R V G
 4641 AATTGGACGG GGTGCGCTA CATAGGTTTG CGCCCCCTG CAAGCCCTG CTGGGGAGG AGGTATCATT CAGAGTAGGA
 TTAACCTGCC CCACGGGAT GTATCCAAAC GCGGGGGAC GTTCGGGAAC GACGCCCTCC TCCATAGTAA GTCTCATCCT

+2 L H E Y P V G S Q L P C E P E P D V A V L T S M L T D
 4721 CTCCACGAAT ACCCGGTAGG GTCGCAATTA CCTTGCAGC CGAACCGGA CGTGGCCGTG TTGACGTCCA TGCTCACTGA
 GAGGTGCTTA TGGCCATCC CAGCGTTAAT GGAACGCTCG GGCTGGCCT GCACCGGCAC AACTGCAGGT ACGAGTGA

+2 P S H I T A E A A G R R L A R G S P P S V A S S S A
 4801 TCCCTCCCAT ATAACAGCAG AGGCGGCCGG CGAAGGTTG GCGAGGGAT CACCCCCCTC TGCGGCCAGC TCCTCGGCTA
 AGGGAGGGTA TATTGTCGTC TCGCCGGCC CGCTTCAAC CGCTCCCTA GTGGGGGAG ACACCGGTG AGGAGCCGAT

pCMV-delNS35

FIGURE 5 - Page 5

+2 S Q L S A P S L K A T C T A N H D S P D A E L I E A N
 4881 GCCAGCTATC CGCTCCATCT CTCAGGCAA CTIGCACCGC TAACCATGAC TCCCCTGATG CTGAGCTCAT AGAGGCCAAC
 CGGTGATAG GCGAGGTAGA GAGTTCCCTT GAACGGCGG ATTGGTACTG AGGGGACTAC GACTCGAGTA TCTCCGGTTG

+2 L L W R Q E M G G N I T R V E S E N K V V I L D S F D
 4961 CTCCCTATGGA GGCAGGAGAT GGGCGGCAAC ATCACCAAGGG TTGAGTCAGA AAACAAAGTG GTGATTCTGG ACTCCCTCGA
 GAGGTACCT CGCTCCTCTA CCCGCCGTG TAGTGGTCCC AACTCAGTCT TTGTTTCAC CACTAGACC TGAGGAAGCT

+2 P L V A E E D E R E I S V P A E I L R K S R R F A Q
 5041 TCCGCTTG GCGGAGGAGG ACGAGCGGA GATCTCGTA CCCGAGAAA TCTGCGGAA GTCTCGGAGA TTGCGCCAGG
 AGGCAACAC CGCCTCTCC TGCTCGCCCT CTAGAGGCAT GGGCGTCTT AGGACGCCCT CAGAGCTCT AAGGGGGTCC

+2 A L P V W A R P D Y N P F L V E T W . K K P D Y E P P V
 5121 CCTGCCCCGT TTGGCGCGG CCGGACTATA ACCCCCCGCT AGTGGAGACG TGAAAAGC CCGACTACGA ACCACCTG
 GGGACGGGCA AACCCGCGCC GGCCTGATAT TGGGGGCGA TCACCTCTGC ACCTTTTCG GGCTGATGCT TGGTGGACAC

+2 V H G C P L P P P K S P P V P P P R K K R T V V L T E
 5201 GTCCATGGCT GCGCGCTTC ACCTCCAAG TCCCTCTG TGCTCCGCGC TCGGAAGAAG CGGACGGTGG TCCCTACTGA
 CAGGTACCGA CGGGCGAAGG TGGAGGTTTC AGGGGAGGAC ACGGAGGCGG AGCCTCTC GCCTGCCACC AGGAGTACT

+2 S T L S T A L A E L A T R S F G S S S T S G I T G D
 5281 ATCAACCTA TCTACTGCCT TGGCGAGCT CGCCACAGA AGCTTGGCA GCTCTCAAC TTCCGGCATT ACGGGGCGACA
 TAGTGGGAT AGATGACGGA ACCGGCTCGA GCGTGGTCT TCGAAACCGT CGAGGAGTTG AAGGCCGTA TGCCCGCTGT

+2 N T T T S S E P A P S G C P P D S D A E S Y S S M P P
 5361 ATACGACAAC ATCCTCTGAG CCCGCCCCCTT CTGGCTGCC CCCCAGCTCC GACGCTGAGT CCTATTCTC CATGGCCCC
 TATGCTGTTG TAGGAGACTC GGGGGGAA GACCGACGGG GGGCGTGAGG CTGCGACTCA GGATAAGGAG GTACGGGGGG

+2 L E G E P G D P D L S O G S W S T V S S E A N A E D V
 BamHI

5441 CTGGAGGGGG AGCCTGGGAA TCCGGATCTT AGCGACGGGT CATGGTCAAC GGTCAAGTAGT GAGGCCAACG CGGAGGATGT
 GACCTCCCCC TCGGACCCCT AGGCCTAGAA TCCCTGCCA GTACCAAGTTG CCAGTCATCA CTCCGGTTGC GCCTCTACA

+2 V C C S M S Y S W T G A L V T P C A A E E Q K L P I
 5521 CGTGTGCTGC TCAATGTCTT ACTCTGGAC AGGGCACTC GTCACCCCGT GCGCCGCGGA AGAACAGAAA CTGCCCCATCA
 GCACACGACG AGTTACAGAA TGAGAACCTG TCCGGTGAG CAGTGGGGCA CGCGGGCCT TCTGTCTT GACGGGTAGT

+2 N A L S N S L L R H H N L V Y S T T S R S A C Q R Q K
 5601 ATGCACTAAG CAACTCGTTC CTACGTCACC ACAATTGGT GTATTCACC ACCTCACGC GTGCTTGCA AAGGCAGAAG
 TACGTGATTC GTTACGCAAC GATGCACTG TGTTAAACCA CATAAGGTGG TGGACTGCGT CACGAACGGT TTCCGTCTTC

+2 K V T F D R L Q V L D S H Y Q D V L K E V K A A A S K
 5681 AAAGTCACAT TTGACAGACT GCAAGTTCTG GACAGCCATT ACCAGGACGT ACTCAAGGAG GTTAAAGCAG CGGCGTCAAA
 TTTCAGTGTAA AACTGTCTGA CGTCAAGAC CTGTCGGTAA TGGTCTGCA TGAGTCCTC CAATTCTGTC CGCGCAGTT

+2 V K A N L L S V E E A C S L T P P H S A K S K F G Y
 5761 AGTGAAGGCT AACTTGCTAT CCGTAGAGGA AGCTTGACG CTGACGCCCC CACACTCAGC CAAATCCAAG TTTGGTTATG
 TCACCTCCGA TTGAACGATA GGCATCTCT TCGAACGTGG GACTGCGGG GTGTGAGTCG GTTAAAGGTC AAACCAATAC

+2 G A K D V R C H A R K A V T H I N S V W K D L L E D N
 5841 GGGCAAAAGA CGTCCGTTGC CATGCCAGAA AGGCCGTAA CCACATCAAC TCCGTGTGGA AAGACCTCTT GGAAGACAAT
 CCCGTTTCT GCAGGCAACG GTACGGTCTT TCCGGCATTG GGTGTAGTTG AGGCACACCT TTCTGGAAGA CCTTCTGTCA

+2 V T P I D T T I M A K N E V F C V Q P E K G G R K P A
 5921 GTAACACCAA TAGACACTAC CATCATGGCT AAGAACGAGG TTTCTGCGT TCAAGCTGAG AAGGGGGGTC GTAAGCCAGC
 CATTGTGGTT ATCTGTGATG GTAGTACCGA TTCTGCTCC AAAAGACGCA AGTCGGACTC TTCCCCCAG CATTGGTCA

pCMV-delNS35

FIGURE 5 - Page 6

+2 R L I V F P D L G V R V C E K M A L Y D V V T K L P
 6001 TCGTCATC GTGTTCCCG ATCTGGCGT CGCGGTGTC GAAAAGATGG CTGGTACGA CGGGTTACA AAGCTCCCT
 AGCAGAGTAG CACAAGGGC TAGACCCGCA CGGCACACG CTTTCTACC GAAACATGCT GCACCAATGTC TTGAGGGGA

+2 L A V M . G S S Y G F Q Y S P G Q R V E F L V Q A W K S
 EcoRI

6081 TGGCCGTGAT GGGAAAGCTCC TACGGATTCC AATACTCACC AGGACAGCGG GTTGAATTCC TCGTCAGTC GTGGAAGTCC
 ACCGGCACTA CCCTTCGAGG ATGCCCTAAGG TTATGAGTGG TCCTGTCGCC CAACTTAAGG AGCACGTTG CACCTTCAGG

+2 K K T P M G F S Y D T R C F D S T V T E S D I R T E E
 6161 AAGAAAACCC CAATGGGGTT CTCGTATGAT ACCCGCTGCT TTGACTCCAC AGTCACTGAG AGCGACATCC GTACGGAGGA
 TTCTTTGGG GTTACCCCCA GAGCATACTA TGGCGACGA AACTGAGGTG TCAGTACTC TCGCTGAGG CATGCCCTCCT

+2 A I Y Q C C D L D P Q A R V A I K S L T E R L Y V G
 6241 GGCAATCTAC CAATGTTGTG ACCTCGACCC CCAAGCCCGC GTGGCCATCA AGTCCTCAC CGAGAGGCTT TATGTTGGGG
 CCGTAGATG GTTACAACAC TGGAGCTGGG GGTCGGGGG CACCGTAGT TCAGGAGTG GCTCTCCGAA ATACAACCCC

+2 G P L T N S R G E N C G Y R R C R A S G V L T T S C G
 6321 CCCCTCTAC CAATTCAARG GGGGAGAAC GCGGCTATCG CAGGTGGCCG CGGAGCGGGG TACTGACAAC TAGCTGTGGT
 CGGGAGAATG GTTAAGTTCC CCCCTCTTGA CGCCGATAGC GTCCACGGCG CGCTCGCCGC ATGACTGTTG ATCGACACCA

+2 N T L T C Y I K A R A A C R A A G L Q D C T M L V C G
 6401 AACACCCCTCA CTTGCTACAT CAAGGCCCGG GCAGCCTGTC GAGCCGAGG GCTCCAGGAC TGCACCATGC TCGTGTGTG
 TTGTGGGAGT GAACGATGTA GTTCCGGGCC CGTCGGACAG CTGGCGTCC CGAGGTCTG ACGTGGTACG AGCACACACC

+2 D D L V V I C E S A G V Q E D A A S L R A F T E A M
 6481 CGACGACTTA GTCGTTATCT GTAAAGCGC GGGGGTCCAG GAGGACGGG CGACCTGAG AGCCTTCAGG GAGGCTATGA
 GCTGCTGAAT CAGCAATAGA CACTTCGCG CCCCCAGGTC CTCTGGCCG GCTCGGACTC TCGGAAGTGC CTCCGATACT

+2 T R Y S A P P G D P P Q P E Y D L E L I T S C S S N V
 6561 CCAGGTACTC CGCCCCCCC GGGGACCCCC CACAACAGA ATACGACTTG GAGCTCATAA CATCATGTC CTCAAACGTG
 GGTCCATGAG GCGGGGGGGG CCCCTGGGGG GTGTTGGTCT TATGCTGAAC CTCGAGTATT GTAGTACGAG GAGGTTGCAC

+2 S V A H D G A G K R V Y Y L T R D P T T P L A R A A W
 6641 TCAGTCGCC ACGACGGCG TGAAAGAGG GTCTACTACC TCACCCGTA CCCTACAACC CCCCTCGCGA GAGCTGCGT
 AGTCAGCGG TGCTGCGCG ACCTTCTCC CAGATGATGG AGTGGCACT GGGATTTGG GGGAGCGCT CTCGACGCAC

+2 E T A R H T P V N S W L G N I I M F A P T L W A R M
 6721 GGAGACAGCA AGACACACTC CAGTCATTC CTGGCTAGGC AACATAATCA TGTTGGCCC CACACTGTGG GCGAGGATGA
 CCTCTGTGAG TCTGTGTGAG GTCAAGTAAAG GACCGATCCG TTGATTAGT ACAACGGGG GTGTGACACC CGCTCCTACT

+2 I L M T H F F S V L I A R D Q L E Q A L D C E I Y G A
 6801 TACTGATGAC CCATTTCTT AGCGTCCTTA TAGCCAGGG CCAGCTGAA CAGCCCTCG ATTGCGAGAT CTACGGGGC
 ATGACTACTG GTAAAGAAA TCGCAGGAAT ATCGGTCCTT GGTCAACTT GTCCGGGAGC TAACGCTCA GATGCCCGG

+2 C Y S I E P L D L P P I I Q R L H G L S A F S L H S Y
 6881 TGCTACTCCA TAGAACCACT GGATCTACCT CCAATCATTC AAAGACTCCA TGGCTCAGC GCATTTTCAC TCCACAGTTA
 ACGATGAGGT ATCTGGTGA CCTAGATGGA GTTGTAGGT TTTCTGAGGT ACCGGAGTCG CGTAAAGTG AGGTGTCAAT

+2 S P G E I N R V A A C L R K L G V P P L R A W R H R
 6961 CTCTCCAGGT GAAATCAATA GGGTGGCCGC ATGCCCTCAGA AAACCTGGGG TACCGCCCTT GCGAGCTGG AGACACCGGG
 GAGAGGTCCA CTTAGTTAT CCCACCGGGC TACGGAGTCT TTTGAACCCC ATGGGGGAA CGCTCGAACCC TCTGTGGCCC

+2 A R S V R A R L L A R G G R A A I C G K Y L F N W A V
 7041 CCCGGAGCGT CGCGCTAGG CTTCTGGCCA GAGGAGCGAG GGCTGCCATA TGTTGGCAAGT ACCTCTCAA CTGGCAGTA
 GGGCTCGCA GGGCGATCC GAAAGACCGGT CTCCCTCCGTC CCGACGGTAT ACACCGTTCA TGGAGAAGTT GACCCGTAT

pCMV-delNS35

FIGURE 5 - Page 7

+2 R T K L K L T P I A A A G Q L D L S G W F T A G Y S G
 7121 AGAACAAAGC TCAAACTCAC TCCAAATAGCG GCGCTGGCC AGCTGGAAC TGTGGCTGG TTACGGCTG CCTACAGCGG
 TCTTGTTCG AGTTGAGTG AGTTATCGC CGGCACCGG TCGACCTGAA CAGGGCGACC AAGTGCCGAC CGATGTCGCC

+2 G D I Y H S V S H A R P R W I W F C L L L L A A G V
 7201 GGGAGACATT TATCACAGCG TGCTCATGC CGGGCCCCCG TGGATCTGGT TTTGCCTACT CCTGCTTGCT GCAGGGGTAG
 CCCTCTGTA ATAGTGTGCG ACAGAGTACG GGCGGGGCG ACCTAGACCA AAACGGATGA GGACGAACGA CGTCCCCATC

+2 G I Y L L P N R
 7281 GCATCTACCT CCTCCCCAAC CGATGAAGGT TGGGGTAAAC ACTCCGGCCT AAAAAAAA AAAAATCTAG AAAGGCGCC
 CGTAGATGGA GGAGGGGTTG GCTACTTCA ACCCCATTG TGAGGCCGA TTTTTTTT TTTTAGATC TTCCGCCG

BamHI MluI

7361 CAAGATATCA AGGATCCACT ACGCGTTAGA GCTCGCTGAT CAGCCTCGAC TGTGCCTCT AGTTGCCAGC CATCTGTTG
 GTTCTATAGT TCCTAGGTGA TGGCAATCT CGAGCGACTA GTCGGAGCTG ACACCGAAGA TCAACGGTCG GTAGACAACA

7441 TTGCCCTCTCC CCCGTGCCTT CTTGACCTT GGAAAGGTGCC ACTCCCACGT TCCTTCCTA ATAAAATGAG GAAATTGCAT
 AACGGGGAGG GGGCACGGAA GGAACCTGGGA CCTTCCACGG TGAGGGTGAC AGGAAAGGAT TATTTACTC CTTAACGTA

7521 CGCATCTGCT GAGTAGGTGT CATTCTATTG TGGGGGTGG GGTGGGGCAG GACAGCAAGG GGGAGGATTG GGAAGACAAT
 GCGTAACAGA CTCATCCACA GTAAGATAAG ACCCCCCACC CCACCCCGTC CTGTCGTTCC CCCCCTAAC CTTCTGTTA

7601 ACGAGGCATG CTGGGGAGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTGCT TCGGCTGCC CGAGCGGTAT
 TCGTCGGTAC GACCCCTCGA GAAGGGCGAG GAGGGAGTGA CTGAGCGACG CGAGCCAGCA AGCCGACGCC GCTGCCATA

7681 CAGCTCACTC AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG
 GTCGAGTGAG TTCCGCCAT TATGCCATA GGTGTCTTAG TCCCCTATTG CGTCCTTCT TGTACACTCG TTTCCGGTC

7761 CAAAGGCCA GGAACCGTAA AAAGGCCGG TTGCTGGCGT TTTCCATAG GCTCCGCC CGTACGAGC ATCACAAAAA
 GTTTCCGGT CCTTGGCATT TTCCGGCGC AACGACCGCA AAAAGGTATC CGAGGCCGGG GGACTGCTCG TAGTGTTTT

7841 TCGACGCTCA AGTCAGAGGT GGCACCCAGC GACAGGACTA TAAAGATACC AGGGCTTTCC CCTGGAAGC TCCCTCGTGC
 AGCTGGAGT TCAGTCTCCA CCGCTTGGG CTGTCCTGAT ATTCTATGG TCCGCAAAGG GGACCTTCG AGGGAGCAGC

7921 GCTCTCTGT TCCGACCCCTG CCGCTTACCG GATACTGTC CGCCTTCTC CCTTCGGAA GCGTGGCGCT TTCTCAATGC
 CGAGAGGACA AGGCTGGAC GGCACATGGC CTATGGACAG GCGGAAAGAG GGAAGCCCTT CGCACCGCGA AAGAGTTACG

8001 TCACGCTGTA GGATCTCAAG TTGCGTGTAG GTCGTICGCT CCAAGCTGGG CTGTTGTCAC GAAACCCCCG TTCAAGCCGA
 AGTGGACAT CCATAGAGTC AAGCCACATC CAGCAAGCGA GGTCGACCC GACACACGTG CTGGGGGCG AAGTCGGGCT

8081 CCGCTGCC TTATCCGGTA ACTATCGTCT TGAGTCAAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG
 GGCACGCCG AATAGGCCAT TGATAGCAGA ACTCAGGGTG GGCATCTG TGCTGAATAG CGGTGACCGT CGTCGGTAC

8161 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTG AAGTGGTGGC CTAACACTGG CTACACTAGA
 CATTGCTCTA ATCGTCTCGC TCCATACATC CGCCACGATG TCTCAAGAAC TTCACCAACG GATTGATGCC GATGTGATCT

8241 AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTA CCTTCGAAA AAGAGTTGGT AGCTCTGAT CGGGAAACAA
 TCCTGTCATA ACCATAGAC GCGAGACGAC TTGGTCAAT GGAAGCTTT TTCTCAACCA TCGAGAACTA GGCGTTTGT

8321 AACCAACGCT GGTAGCGGTG GTTTTTGT TTGCRAGCAG CAGATTACGC GCAGAAAAA AGGATCTCAA GAAGATCCTT
 TTGGTGGCGA CCATGCCAC CAAAAAAACA AACGTTGTC GTCTAATGCG CGTCTTTTT TTCTAGAGT CTTCTAGGAA

8401 TGATCTTTTACGGGGTCT GACGCTCAGT GGAACGAAA CTCACGTTAA GGGATTTGG TCATGAGATT ATCAAAAGG
 ACTAGAAAAG ATGCCAGA CTGCGAGTCA CCTGCTTT GAGTGCATT CCCTAAAACC AGTACTCTAA TAGTTTCC

pCMV-delNS35

FIGURE 5 - Page 8

8481 ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTA AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG
TAGAAGTGGA TCTAGGAAA TTAAATTCTT ACTTCAATAT TTAGTTAGAT TTCAATATATA CTCATTGAA CCAGACTGTC

8561 TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCATTTT GTTCATCCAT AGTTGCCGTA CTCCCCGTCG
AATGGTTACG AATTAGTCAC TCCGTGGATA GAGTCGCTAG ACAGATAAG CAACTAGGTA TCAACGGACT GAGGGGCAGC

8641 TGTTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGTGCA ATGATAACCGC GAGACCCACG CTCACCGGCT
ACATCTATTG ATGCTATGCC CTCCCGAATG GTAGACCGGG GTCACGACGT TACTATGGCG CTCTGGGTGC GAGTGGCCGA

8721 CCAGATTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGTCCTGCA ACTTTATCCG CCTCCATCCA
GGTCTAAATA GTCGTTATTG GTCGGTCCG CCTTCCCGC TCGGTCTTC ACCAGGACGT TGAAATAGGC GGAGGTAGGT

8801 GTCTTAAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTGCCTCAA CGTTGTTGCC ATTGCTACAG
CAGATAATTA ACAACGGCCC TTCGATCTCA TTCATCAAGC GGTCAATTAT CAAACCGGTT GCAACAACGG TAACGATGTC

8881 GCATCGTGGT GTCACGCTCG TCGTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCAGT TACATGATCC
CGTAGCACCA CAGTGCAGC AGCAACCAT ACCGAAGTAA GTCGAGGCCA AGGGTGCTA GTTCCGCTCA ATGTTACTAGG

8961 CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTGGCTCCTC CGATCGTGT CAGAAGTAAG TTGGCCGAG TGTTATCACT
GGTACAACA CGTTTTTCG CCAATCGAGG AAGCCAGGAG GCTAGCAACA GTCTTCATTC AACGGCGTC ACAATAGTGA

9041 CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCCGTAATGATGTTTTTG TGTCAGTGGT GAGTACTCAA
GTACCAATAC CGTCGTGACG TATTAAGAGA ATGACAGTAC GGTAGGCATT CTACGAAAG ACACGTGACCA CTCATGAGTT

9121 CCAAGTCATT CTGAGAATAG TGATGCGGC GACCGAGTTG CTCTGCCCCG GCGTCATAC GGATAATAC CGCGCCACAT
GGTCAGTAA GACTCTTATC ACATACGCCG CTGGCTAAC GAGAACGGGC CGCAGTTATG CCCTATTATG GCGCGGTGTA

9201 ACCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTCTT CGGGGGGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC
TCCTCTTGAATTTTACGA GTAGTAACCT TTGCAAGAA GCCCCGTTTG TGAGAGTTCC TAGAATGGCG ACAACTCTAG

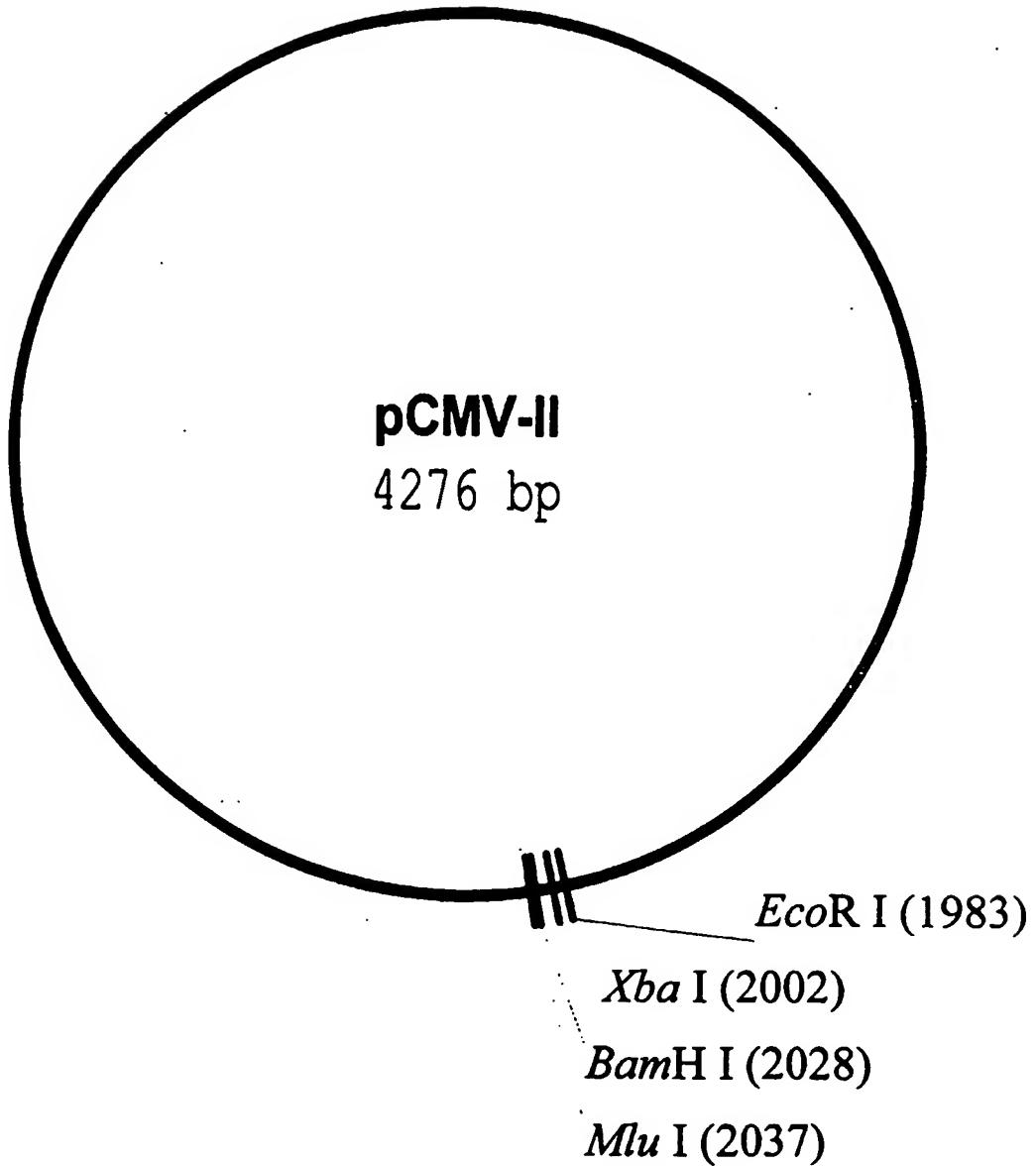
9281 CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTA CTTTCACCAAG CGTTCTGGG TGAGCAAAA
GTCAAGCTAC ATTGGGTGAG CACGTGGTT GACTAGAAAGT CGTAGAAAT GAAAGTGGTC GCAAGAACCC ACTCGTTTT

9361 CAGGAAGGCA AAATGCCGA AAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCTT TTTCAATAT
GTCCTTCCGT TTTACGGGT TTTTCCCTT ATTCCCGCTG TGCCCTTACA ACTTATGAGT ATGAGAAGGA AAAAGTTATA

9441 TATTGAAGCA TTATCAGGG TTATTGTCCTC ATGAGCGGAT ACATATTGA ATGTATTTAG AAAATAAAC AAATAGGGT
ATAACTTCGT AAATAGTCCC AATAACAGAG TACTGCCCTA TGATAAAACT TACATAARTC TTTTATTTG TTATCCCCA

9521 TCCGCGCACA TTTCCCGAA AAGTGCCACC TGACGTCTAA GAAACCATTA TTATCATGAC ATTAACCTAT AAAATAGGC
AGGGCGTGT AAAGGGGTTT TTCACGGTGG ACTGCAGATT CTTGGTAAT ATAGTACTG TAATTGGATA TTTTATCCG

9601 GTATCACGAG GCCCTTTCGT C
CATAGTGCCTC CGGGAAAGCA G

FIGURE 6

pCMV-II

FIGURE 7 - Page 1

1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCITGTCT GTAAGCGGAT
AGCGGCCAAA GCCACTACTG CCACCTTTGG AGACTGTGTA CGTCGAGGGC CTCTGCCAGT GTCGAACAGA CATTGCCATA

81 GCGGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA
CGGCCCTCGT CTGTTGGGC AGTCCCAGC AACGGCCAC AGCCCCGACC GAATTGATAC GCCGTAGTCT

161 GCAGATTGTA CTGAGAGTGC ACCATATGAA GCTTTTGCA AAAGCCTAGG CCTCCAAAAA AGCCTCCICA CTACTCTGG
CGTCTAACAT GACTCTCACG TGGTATACTT CGAAAACGT TTTCGGATCC GGAGGTTTT TCAGGAGGACT GATGAAGACC

241 AATAGCTAG AGGGCGAGGC GGCTCGGCC TCTGCATAAA TAAAAAAAAT TAGTCAGCCA TGGGGCGGAG AATGGGCCA
TTATCGAGTC TCCGGCTCCG CCGGAGCCGG AGACGTATT ATTTCCTTA ATCAGTCGGT ACCCCGCCTC TTACCGCCT

321 ACTGGCGGG QAGGGAATTA TTGGCTATTG GCCATTGCA ACAGTTGATC TATATCATAA TATGTACATT TATATTGGCT
TGACCCGCC CTCCCTTAAT AACCGATAAC CGGTAACCTA TGCAACATAG ATATAGTATT ATACATGAA ATATAACCGA

401 CATGTCCAAT ATGACCGCA TGTTGACATT GATTATTGAC TAGTTATTA TAGTAACTAA TTACGGGGTC ATTAGTCAT
GTACAGGTTA TACTGGCGT ACAACTGTAA CTAATAACTG ATCAATAATT ATCATTAGTT ATAGCCCCAG TAATCAAGTA

481 AGCCCATATA TGGAGTTCCG CGTTACATAA CCTACGGTAA ATGGCCGCC TGGCTGACCG CCCACGACCC CGGCCATT
TCGGGTATAT ACCTCAAGGC GCAATGTATT GAATGCCATT TACCGGGCG ACCGACTGGC GGGTGTGG GGGCGGGTAA

561 GACGTCAATA ATGACGTATG TTCCCAGTAAACGCCATAA GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTACGGT
CTGAGTTAT TACTGCATAC AAGGGTATCA TTGGGTAT CCTGAAAGG TAATGCACTG TACCCACCTC ATAAATGCCA

641 AAACGTCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTCCGCC CCTATTGACCG TCAATGACGG TAAATGGCC
TTTGACGGGT GAACCGTCAT GTAGTTACAA TAGTATAACGG TTCAGGGGG GGATAACTGC AGTTACTGCC ATTTACCGG

721 GCCTGGCATT ATGCCAGTA CATGACCTTA CGGGACTTC CTACTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC
CGGACCGTAA TACGGGTATC GTACTGGAAT GCCCTGAAAG GATGAAACGT CATGAGATG CATAATCAGT AGCGATAATG

801 CATGGTGATG CGGTTTGGC AGTACACCAA TGGCGTGGAA TAGCGGTITG ACTCACGGG ATTCCAAGT CTCCACCCCA
GTACCAACTAC GCCAAAACCG TCATGTGGTT ACCCGCACCT ATCGCCAAAC TGAGTGGCC TAAAGGTTCA GAGGTGGGT

881 TTGACGTCAA TGGGAGTTG TTTGGCACC AAAATCAACG GGACTTTCA AAATGCGTA ATAACCCCGC CCCGTTGACCG
AACTGCAGTT ACCCTCAAAC AAAACCGTGG TTTAGTTGC CCTGAAAGGT TTACAGCAT TATTGGGCG GGGCAACTGC

961 CAAATGGCGG GTAGGGGTGT ACGGTGGAG GTCTATATAA GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG
GTTTACCCGC CATCCGCACA TGCCACCCCTC CAGATATATT CGTCTCGAGC AAATCACTG GCAGTCTAGC GGACCTCTGC

1041 CCATCCACGC TGTGTTGACC TCCATAGAAG ACACCGGAC CGATCCAGCC TCCGCGGCCG GGAACGGTGC ATTGGAACCC
GGTAGGTGGC ACAAAACTGG AGGTATCTTC TGTGGCCCTG GCTAGGTGG AGGCGCCGGC CCTTGCCACG TAACTTGCG

1121 GGATTCCCCG TGCCAAGAGT GACGTAAGTA CCGCCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA
CCTAAGGGGC ACAGTTCTCA CTGCATTATC GGCGATATC TGAGATATCC GTGTGGGAA ACCGAGAATA CGTACGATAT

1201 CTGTTTTGG CTGGGGCCT ATACACCCCC GCTCCTTATG CTATAGGTGA TGGTATAGCT TAGCCTATAG GTGTGGGTTA
GACAAAAACG GAACCCCGGA TATGTGGGG CGAGGAATAC GATATCCACT ACCATATCGA ATCGGATATC CACACCCAA

1281 TTGACCATTA TTGACCACTC CCCTATTGGT GACGATATT TCCATTACTA ATCCATAACA TGGCTCTTG CCACAACTAT
AACTGTAAT AACTCGTGAG GGGATAACCA CTGCTATGAA AGGTAATGAT TAGTATTGT ACCGAGAAGC GGTGTTGATA

1361 CTCTATTGGC TATATGCCAA TACTCTGTCC TTCAGAGACT GACACGGACT CTGTATTTT ACAGGATGGG GTCCATTAT
GAGATAACCG ATATACGGTT ATGAGACAGG AAGTCTCTGA CTGTCCTGA GACATAAAA TGTCCCTACCC CAGTAAATA

1441 TATTTCACAA TTCACATATA CAACAAACGCC GTCCCCCGTG CCCGAGTAA TATTAACCA TAGCGTGGGA TCTCCGACAT
ATAATGTTT AAGTGTATAT GTTGTGCGG CAGGGGGAC GGGCGTCAA AATAATTGT ATCGCACCC AGAGGCTGTA

pCMV-II

FIGURE 7 - Page 2

1521 CTCGGGTACG TGTCGGAC ATGGGCTCTT CTCCGGTAGC GGCGGAGCTT CCACATCCGA GCCCTGGTCC CATCCGCCA
GAGCCCATGC ACAAGGCCTG TACCCGAGAA GAGGCATCG CCGCCTGAA GGTGTAGGCT CGGGACCGG GTAGGCAGGT

1601 GCGGCTCATG GTCCGCTGGC AGCTCCTTGC TCCTAACAGT GGAGGCCAGA CTTAGGCACA GCACAATGCC CACCACCA
CGCCGACTAC CAGCGAGCCG TCGAGGAACG AGGATTGTC CCTCCGGTCT GAATCCGTGT CGTGTACGG GTGGTGGTGG

1681 AGTGTGCCGC ACAAGGCCGT GGCGGTAGGG TATGTGTCTG AAAATGAGCT CGGAGATTGG GCTCGCACCT GGACGCAGAT
TCACACGGCG TGTCGGCA CGGCCATCCC ATACACAGAC TTTTACTCGA CCTCTAACCG CGAGCGTGGA CCTCGCTCA

1761 GGAAGACTTA AGGCAAGGGC AGAAGAAGAT GCAGGCAGCT GAGTTGTGT ATTCTGATAA GAGTCAGAGG TAACTCCCGT
CCTCTGAAT TCCGTCGGCG TCTTCTCTA CGTCCGGTCA CTCAACRACA TAAGACTATT CTAGTCTCC ATTGAGGGCA

1841 TGCGGTGCTG TAAACGGTGG AGGGCAGTGT AGTCTGAGCA GTACTCGTTG CTGCCGCGCG CGCCACCCAGA CATAATAGCT
ACGCCACCG AATTGCCACC TCCGTCACA TCAGACTCGT CATGACCAAC GACGGCGCGC GCGGTGGTCT GTATTATCGA

EcoRI

1921 GACAGACTAA CAGACTGTTC CTTCCATGG GTCTTTCTG CAGTCACCGT CGTCGACCTA AGAATTCAAGA CTCGAGCAAG
CTGTCTGATT GTCTGACAAG GAAAGGTACC CAGAAAAGAC GTCACTGGCA GCAGCTGGAT TCTTAAGTCT GAGCTCGTTC

XbaI

BamHI

MluI

2001 TCTAGAAAG CGCGCCAAGA TATCAAGGAT CCACTACCGG TTAGAGCTCG CTGATCAGCC TCGACTGTGC CTCTAGTTG
AGATCTTCC CGCCGGTTCT ATAGTTCCTA GGTGATCGC AATCTCGAGC GACTACTCGG AGCTGACACG GAAGATCAAC

2081 CCAGCCATCT GTTGTGGCC CCTCCCCCGT GCCTTCCCTG ACCCTGGAAG GTGCCACTCC CACTGTCCCTT TCCTAATAAA
GGTCGGTAGA CAACAAACGG GGAGGGGGCA CGGAAGGAAC TGGGACCTTC CACGGTGAGG GTGACAGGAA AGGATTATTT

2161 ATGAGGAAAT TGCATCGAT TGTCTGAGTA GGTGTCAATT TATTCTGGGG GGTGGGGTGG GGCAAGGACAG CAAGGGGGAG
TACTCCTTA ACGTAGCGTA ACAGACTCAT CCACAGTAAG ATAAGACCCC CCACCCACC CCGTCTGTC GTTCCCCCTC

2241 GATTGGGAAG ACAATAGCAG GCATGCTGGG GAGCTCTTCC GCTTCCCGC TCACTGACTC GCTGCGCTCG GTCGTTCGGC
CTAACCCCTTC TGTATCGTC CGTACGACCC CTCGAGAAGG CGAAGGAGCG AGTGAATGAG CGACCGGAGC CAGCAAGCC

2321 TGCGCGAGC GGATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGAGG AAAGAACATG
ACGGCGCTCG CCATAGTCGA GTGAGTTCC GCCATTATGC CAATAGGTGT CTTAGTCCCC TATTGCGTCC TTTCTGTAC

2401 TGACGAAAG GCCAGAAAA GGCCAGGAAC CGTAAAAAGG CGCGGTGCT GGCCTTTTC CATAGGCTCC GCCCCCTGA
ACTCGTTTC CGGTGTTT CCGGTCTTG GCATTTTCC GCGCAACGA CGCAAAAG GTATCCGAGG CGGGGGACT

2481 CGAGCATCAC AAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCGACAG GACTATAAG ATACCAAGGCG TTTCCCCCTG
GCTCGTAGTG TTTTAGCTG CGAGTTCACT CTCCACCGCT TTGGGCTGTC CTGATATTC TATGGTCCCG AAAGGGGGAC

2561 GAAGCTCCCT CGTGCCTCT CCTGTTCCGA CCCTGCCGT TACCGGATAC CTGTCGCCCT TTCTCCCTC GGGAGCGTG
CTTCGAGGGGA GCACCGGAGA GGACAAGGCT GGGACGGGA ATGGCTATG GACAGGCGGA AAGAGGGAA CGCTTCGAC

2641 GCGCTTCTC AATGCTCACG CTGTAGGTAT CTCACTCGG TGAGGTGCTG TCGCTCCAAG CTGGGCTGTG TGACGAACC
CGCGAAAGAG TTACGAGTGC GACATCCATA GAGTCAGGCC ACATCCAGCA AGCGAGGTTG GACCCGACAC ACGTGCTTGG

2721 CCCCCCTTCAG CCCGACCGCT CGGCCTTATC CGGTAACAT CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATGCCAC
GGGGCAAGTC CGGCTGGCGA CGCGGAATAG GCCATGATA GCAGACTCA GTTGGGCGA TTCTGTGCTG AATACGGTG

2801 TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGGGTA TGAGGGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC
ACCGTCGTCG GTGACCATGG TCCTAATCGT CTCGCTCCAT ACATCCGCA CGATGTCTCA AGAACCTCAC CACCGGATTG

2881 TACGGCTACA CTAGAAGGAC AGTATTTGGT ATCTGCGTC TGCTGAAGCC AGTACCTTC GGAAAAAGAG TTGGTAGCTC
ATGCCGATGT GATCTCCCTG TCATAAACCA TAGACCGAG ACGACTCGG TCAATGGAAG CCTTTCTC AACCATCGAG

pCMV - II

FIGURE 7 - Page 3

2961 TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTT TTGTTTGCA AGCAGCAGAT TACGCCAGA AAAAAAGGAT
AACTAGGCCG TTGTTGGT GGCGACCATC GCCACCAAA AAACAAACGT TCGTCGCTA ATGCGCGTCT TTTTTCCCTA

3041 CTCAGAAGA TCCTTGATC TTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTCAC CTTAAGGGAT TTGTCATG
GAGTCTTCT AGGAAACTAG AAAAGATGCC CCAGACTGGG AGTCACCTG CTTTGAGTG CAATTCCCTA AAACCACTAC

3121 AGATTATCAA AAAGGATCTT CACCTAGATC CTTTAATT AAAATGAAG TTTAAATCA ATCTAAAGTA TATATGAGTA
TCTAATAGTT TTCTAGAA GTGGATCTAG GAAAATTAA TTTTACTTC AAAATTAGT TAGATTCTAT ATATACTCAT

3201 AACCTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTAG CGATCTGTCT ATTCGTTCA TCCATAGTTG
TTGACCCAGA CTGTCAATGG TTACGAATTA GTCACTCCGT GGATAGAGTC GCTAGACAGA TAAAGCAAGT AGGTATCAAC

3281 CCTGACTCCC CGTCGTGAG ATAACATCGA TACGGGAGGG CTTACCATCT GGCCCCAGTG CTGCAATGAT ACCGGAGAC
GGACTGAGGG GCAGCACATC TATTGATGCT ATGCCCTCCC GAATGGTACA CGGGGTCAC GACGTTACTA TGGCGCTCTG

3361 CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCCAGC CAGCCGAAG GGCCGAGCG AGAAGTGGTC CTGCAACTT
GGTCCGAGTG CCCGAGGTCT AAATAGTCGT TATTGGTCG GTCGGCCTTC CGGCGTCG TCTTACCCAG GACGTTGAAA

3441 ATCCGCCTCC ATCCAGTCTA TTAATTGTT CGGGGAAGCT AGAGTAAGTA GTTCGCACT TAATAGTTG CGCAACGTTG
TAGGCGGAGG TAGGTCAAGAT AATTAACAC GGCCTTCGA TCTCATTCTAT CAACGGTCA ATTATCAAAC GCGTTGCAAC

3521 TTGCCATTGC TACAGGCATC GTGGTGTAC GCTCGTCGTT TGATGGCT TCATTCAGCT CCGGTTCCCA ACCATCAAGG
AACGGTAACG ATGTCGTAG CACCACTG CGAGCACAA ACCATACCGA AGTAAGTCGA GGCCAGGGT TGCTAGTTCC

3601 CGAGTTACAT GATCCCCCAT GTGTGCAAA AAAGCGGTTA GCTCCCTCGG TCCTCCGATC GTGTCAGAA GTAAGTTGGC
GCTCAATGTA CTAGGGGTA CAACACGTTT TTGCAAT CGAGGAAGCC AGGAGGCTAG CAACAGTCTT CATTCAACCG

3681 CGCAGTGTAA TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTACTG TCATGCCATC CGTAAGATGC TTTCTGTGA
GCGTCACAAT AGTGAGTACCA AATACCGTCG TGACGTATTA AGAGAATGAC AGTACGGTAG GCATTCTACG AAAAGACACT

3761 CTGGTAGTA CTCAACCAAG TCATTCTGAG AATAGTGTAT GCGGGGACCG AGTGCTCTT GCGGGCGTC AATACGGGAT
GACCACTCAT GAGTGGTTC AGTAAGACTC TTATCACATA CGGCGTCGG TCAACGAGAA CGGGCCGAG TTATGCCCTA

3841 AATACCGGC CACATAGCAG AACTTTAAA GTGCTCATCA TTGGAAAAGG TTCTTCGGGG CGAAAACCTCT CAAGGATCTT
TTATGGCCG GTGTATCGTC TTGAAATTTC CACGAGTGT AACCTTTGC AAGAAGCCCC GCTTTGAGA GTTCCTAGAA

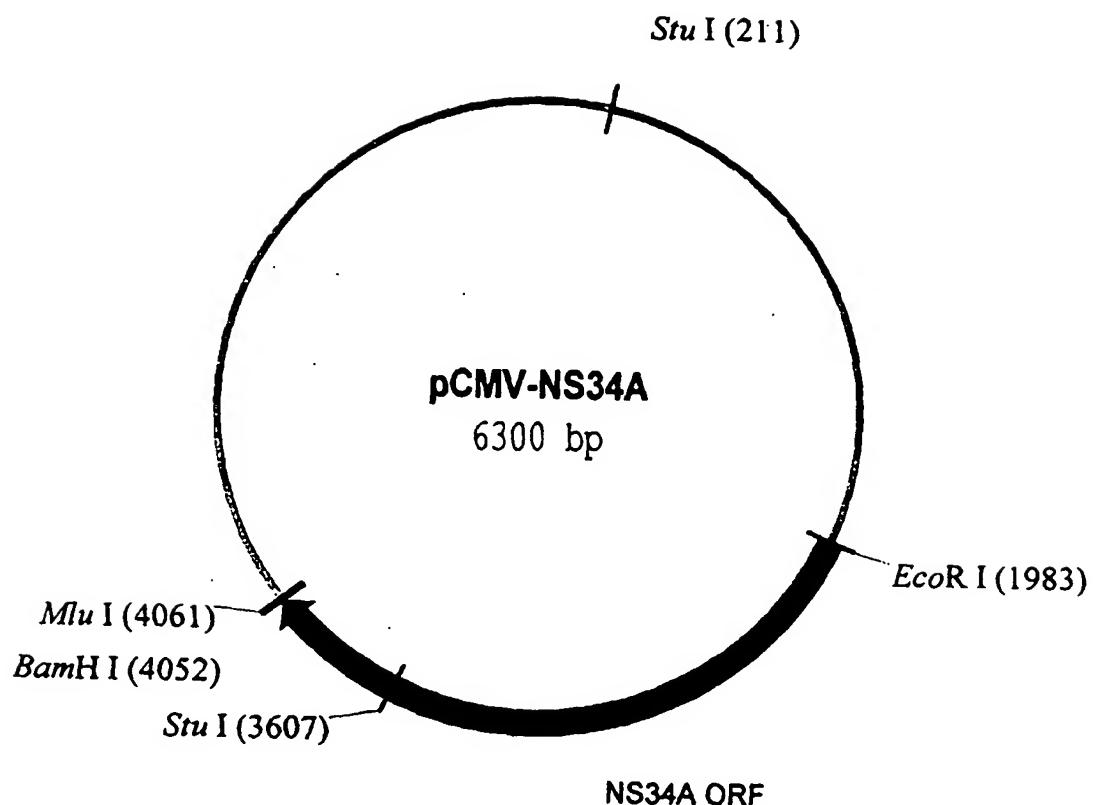
3921 ACCGCTGTG AGATCCAGTT CGATGTAACC CACTCGTCA CCCAAGTGT CTTCAGCATC TTTTACTTTC ACCAGCGTT
TGGCGACAAC TCTAGGTCAA GCTACATTGG GTGAGCACGT GGGTTGACTA GAAGTCGTAG AAAATGAAAG TGGTCGCAAA

4001 CTGGGTGAGC AAAACAGGA AGGCAAAATG CGCAAAAAA GGGATAAGG GCGACACCGA AATGTTGAAT ACTCATACTC
GACCCACTCG TTTTGTCTT TCCGTTTAC GGCGTTTTT CCCTTATTCC CGCTGTGCCT TTACAACCTA TGAGTATGAG

4081 TTCTTTTC AATATTATTG AAGCATTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA
AAGAAAAAG TTATAATAAC TTGCTAAATA GTCCAATAA CAGAGTACTC GCCTATGTAT AAATCTTTT

4161 TAAACAAATA GGGGTCCCG GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAC CATTATTATC ATGACATTAA
ATTIGTTAT CCCAAGGGC CGTGTAAAGG GGCTTTAC GGTGGACTGC AGATTCTTG GTAATAATAG TACTGTAATT

4241 CCTATAAAA TAGGCGTATC ACGAGGCCCT TTGTC
GGATATTTT ATCCGATAG TGCTCCGGAA AAGCAG

FIGURE 8

pCMV-NS34A

FIGURE 9 - Page 1

TCGGCGGT TT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
 AGCGCGAAA GCCACTACTG CCACITTTGG AGACTGTGTA CGTCGAGGGC

51 GAGACGGTCA CAGCTTGCT GTAAAGCGGAT GCCGGGAGCA GACAAGCCCG
 CTCTGCCAGT GTCGAACAGA CATTGCGCTA CGGCCCTCGT CTGTTGGGC

101 TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
 AGTCCCAGC AGTCGCCAC AACCGCCAC AGCCCCGACC GAATTGATAAC

151 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGAA CCTTTTGCA
 GCGCTAGTCT CGCTAACAT GACTCTCACG TGGTATACTT CGAAAAACGT

StuI

201 AAAGCCTAGG CCTCCAAAAAA AGCCTCCCTCA CTACTTCCTGG AATAGCTCAG
 TTTCGGATCC GGAGGTTTT TCAGGAGGAGT GATGAAGACC TTATCGAGTC

251 AGGCGGAGGC GCCCTCGGCC TCTGCATAAA TAAAAAAAAT TAGTCAGCCA
 TCCGGCTCCG CCGGAGCCCG AGACGTATT ATTTCCTTA ATCAGTCGGT

301 TGGGGCGGAG AATGGGGCGGA ACTGGGCGGG GAGGGAAATA TTGGCTATTG
 ACCCCGCTC TTACCCGCCCT TGACCCGCCCT CTCCCTTAAT AACCGATAAC

351 GCCATTGCAT ACGTGTATC TATATCATAA TATGTACATT TATATTGGCT
 CGGTAACTGTA TCGAACATAG ATATAGTATT ATACATGAA ATATAACCGA

401 CATGTCCAAT ATGACCGCCA TGGTGACATT GATTATTGAC TAGTTATTAA
 GTACAGGTTA TACTGGCGGT ACAACTGTAA CTAATAACTG ATCAATAATT

451 TAGTAATCAA TTACGGGGTC ATTAGTTCAT ACCCCATATA TGGAGTICCG
 ATCATTAGTT AATGCCCGAG TAATCAAGTA TCGGGTATAT ACCTCAAGGC

501 CGTTACATAA CTACGGTAA ATGGCCGCC TGGCTGACCG CCCAACGACC
 GCAATGTATT GAATGCCATT TACCGGGCGG ACCGACTGGC GGGTTGCTGG

551 CCCGCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA
 GGGCGGTAA CTGCAGTTAT TACTGCATAC AAGGGTATCA TTGCGGTTAT

601 GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AACTGCCCC
 CCCTGAAAGG TAACTGCAGT TACCCACCTC ATAAATGCCA TTTGACGGGT

651 CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTCCGCC CCTATTGACG
 GAACCGTCAT GTAGTTACACA TAGTATACGG TTCAGGGGG GGATAACTGC

701 TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTA
 AGTTACTGCC ATTACCGGG CGGACCGTAA TACGGGTCA GTACTGGAAT

751 CGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC
 GCCCTGAAAG GATGAACCGT CATGTAGATG CATAATCACT AGCGATAATG

801 CATGGTGTG CGGTTTGGC AGTACACCAA TGGGCGTGG AAGCGGTTG
 GTACCACTAC GCAAAACCG TCATGTGGTT ACCCGCACCT ATCGCCAAC

851 ACTCACGGGG ATTCCAAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTG
 TGAGTCCCC TAAAGGTTCA GAGGTGGGGT AACTGCAGTT ACCCTCAAAC

pCMV-NS34A

FIGURE 9 - Page 2

901 TTTTGGCACC AAAATCAACG GGACTTCCA AAATGTCGTA ATAACCCCGC
AAAACCGTGG TTTTAGTTGC CCTGAAAGGT TTACAGCAT TATTGGGCG

951 CCCGTTGACG CAAATGGGCG GTAGGGCGTGT ACGGTGGGAG GTCTATATAA
GGGCAACTGC GTTACCCCGC CATCCGCACA TGCCACCCCTC CAGATATATT

1001 GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC
CGTCTCGAGC AAATCACTTG GCAGTCTAGC GGACCTCTGC GGTAGGTGCG

1051 TGTGGGACCC TCCATAGAACG ACACCGGGAC CGATCCAGCC TCCGGGGCCG
ACAAAAACTGG AGGTATCTTC TGTGGCCCTG GCTAGGTGCG AGGGCCCGC

1101 GGAACGGTGC ATTGGAACGC GGATTCCCCG TGCCAAGAGT GACGTAAGTA
CCTGCCACG TAACCTTGGG CCTAAGGGC ACGGTTCTCA CTGCATTAT

1151 CCGCCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA
GGCGGATATC TGAGATATCC GTGTGGGAA ACCGAGAATA CGTACGATAT

1201 CTGTTTTGG CTTGGGGCCT ATACACCCCC GCTCCTTATG CTATAGGTGA
GACAAAAAACC GAACCCCGA TATGTGGGG CGAGGAATAC GATATCCACT

1251 TGGTATAGCT TAGCCTATAG GTGTGGGTTA TTGACCATTA TTGACCACTC
ACCATATCGA ATCGGATATC CACACCCAAT AACTGGTAAT AACTGGTGAG

1301 CCCTATTGGT GACGATACTT TCCATTACTA ATCCATAACA TGGCTCTTG
GGGATAARCCA CTGCTATGAA AGGTAATGAT TAGGTATTGT ACCGAGAAC

1351 CCACAACTAT CTCTATTGGC TATATGCCAA TACTCTGTCC TTCAGAGACT
GGTGTGATA GAGATAACCG ATATAACGTT ATGAGACAGG AAGTCTCTGA

1401 GACACGGACT CTGTATTTT ACAGGATGGG GTCCATTAT TATTTACAAA
CTGTGCCTGA GACATAAAA TGTCTACCC CAGGTAATAA ATAAATGTTT

1451 TTACATATA CAACAACGCC GTCCCCCGTG CCCGCAGTTT TTATTAACA
AAGTGTATAT GTTGTGCGG CAGGGGCAC GGGGTCAAA ATAATTTGT

1501 TAGCGTGGGA TCTCCGACAT CTGGGTACG TGTTCCGGAC ATGGGCTCTT
ATCGCACCCCT AGAGGCTGTA GAGCCCATGC ACAAGGCCTG TACCCGAGAA

1551 CTCCGGTAGC GGCGGAGCTT CCACATCCGA GCCCTGGTCC CATCCGTCCA
GAGGCCATCG CCGCCCTCGAA GGTGTAGGCT CGGGACCAAG GTAGGCAGGT

1601 GCGGCTCATG GTCGCTCGGC AGCTCCTTGC TCCTAACAGT GGAGGCCAGA
CGCCGAGTAC CAGCGAGCCG TCGAGGAACG AGGATTGTCA CCTCCGGTCT

1651 CTTAGGCACA GCACAATGCC CACCACCAAGTGTGCCG ACAAGGCCGT
GAATCCGTGT CGTGTACGG GTGGTGGTGG TCACACGGCG TGTTCCGGCA

1701 GCGGGTAGGG TATGTGTCTG AAAATGAGCT CGGAGATTGG GCTCGCACCT
CCGCCATCCC ATACACAGAC TTTTACTCGA GCCTCTAACCGAGCGTGG

1751 GGACGCAGAT GGAAGACTTA AGGCAGCGGC AGAAGAAGAT GCAGGCAGCT
CCTCGTCTA CCTTCTGAAT TCCGTGCCG TCTTCTTCTA CGTCCGTCGA

1801 GAGTTGTTGT ATTCTGATAA GAGTCAGAGG TAACCTCCGT TGCAGGTGCTG
CTCAACAACA TAAGACTATT CTCAGTCTCC ATTGAGGGCA ACGCCACGAC

pCMV-NS34A

FIGURE 9 - Page 3

1851 TTAACGGTGG AGGGCAGTGT AGTCTGAGCA GTACTCGTTG CTGCCGGCG
AATTGCCACC TCCCCTCACCA TCAGACTCGT CATGAGCAAC GACGGCGCGC

1901 CGCCACCAAGA CATAATAGCT GACAGACTAA CAGACTGTTG CTTTCATGG
GCGGTGGTCT GTATTATCGA CTGTCATGATT GTCTGACAAG GAAAGGTAC

+2

M A P

EcoRI

1951 GTCTTTCTG CAGTCACCGT CGTCGACCTA AGAATTCCACC ATGGCGGCCA
CAGAAAAGAC GTCAAGTGGCA GCAGCTGGAT TCTTAAGTGG TACCGCGGGT

+2 I T A Y A Q Q T R G L L G C I I T
2001 TCACGGCGTA CGCCCCAGCAG ACAAGGGGGC TCCTAGGGTG CATAATCACC
AGTGGCGCAT GCGGGTCGTC TGTTCCCCGG AGGATCCCAC GTATTAGTGG

+2 S L T G R D K N Q V E G E V Q I V
2051 AGCCTAACTG GCCGGGACAA AAACCAAGTG GAGGGTGAGG TCCAGATTGT
TCGGATTGAC CGGCCCTGTT TTGGTTCAC CTCCCACCTCC AGGTCTAAC

+2 S T A A Q T F L A T C I N G V C
2101 GTCAACTGCT GCCCAAACCT TCCTGGCAAC GTGCATCAAT GGGGTGTGCT
CAGTTGACGA CGGGTTTGGA AGGACCGTTG CACGTAGTTA CCCCACACGA

+2 W T V Y H G A G T R T I A S P K G
2151 GGACTGTCTA CCACGGGGCC GGAAACGAGGA CCATCGCGTC ACCCAAGGGT
CCTGACAGAT GGTGCCCCGG CCTTGCTCCT GTAGCGCAG TGGTTCCCCA

+2 P V I Q M Y T N V D Q D L V G W P
2201 CCTGTCTACCC AGATGTATAC CAATGTAGAC CAAGACCTTG TGGCTGGCC
GGACAGTAGG TCTACATATG GTTACATCTG GTTCTGAAAC ACCCGACCGG

+2 A S Q G T R S L T P C T C G S S
2251 CGCTTCGCAA GGTACCCGCT CATTGACACC CTGCACTTGC GGCTCCTCGG
GCGAAGCGTT CCATGGGCGA GTAACTGTGG GACGTGAACG CCGAGGAGCC

+2 D L Y L V T R H A D V I P V R R K
2301 ACCTTACCT GGTACAGGAGG CACGCCGATG TCATTCCCGT GCGCCGGCGG
TGGAAATGGA CCAGTGTCTCC GTGCGGCTAC AGTAAGGGCA CGCGCCCGCC

+2 G D S R G S L L S P R P I S Y L K
2351 GGTGATAGCA GGGGCAGCCT GCTGTCGCC CGGCCCATTT CCTACTTGAA
CCACTATCGT CCCCCGTGGA CGACAGCGGG GCGGGTAAA GGATGAAC

+2 G S S G G P L L C P A G H A V G
2401 AGGCTCTCG GGGGGTCCGC TGTTGTGCC CGGGGGCAC GCGCTGGCA
TCCGAGGAGC CCCCCAGGCG AACACACGGG GCGCCCCGTG CGGCACCCGT

+2 I F R A A V C T R G V A K A V D F
2451 TATTTAGGGC CGCGGTGTGC ACCCGTGGAG TGGCTAAGGC GGTGGACTTT
ATAAATCCCG CGGCCACACG TGGGCACCTC ACCGATCCG CCACCTGAAA

+2 I P V E N L E T T M R S P V F T D
2501 ATCCCTGTGG AGAACCTAGA GACAACCATG AGGTCCCCGG TGTTCACGG
TAGGGACACC TCTTGGATCT CTGTTGGTAC TCCAGGGGCC ACAAGTGCCT

pCMV-NS34A

FIGURE 9 - Page 4

+2 N S S P P V V P Q S F Q V A H L
 2551 TAACTCCTCT CCACCAAGTAG TGCCCCAGAG CTTCCAGGTG GCTCACCTCC
 ATTGAGGAGA GGTGGTCATC ACGGGGTCTC GAAGGTCCAC CGAGTGGAGG

+2 H A P T G S G K S T K V P A A Y A
 2601 ATGCTCCAC AGGCAGCGGC AAAAGCACCA AGGTCCCAGC TGCAATATGCA
 TAGGAGGTG TCCGTCGCCG TTTCTGGT TCCAGGGCCG ACGTATACTG

+2 A Q G Y K V L V L N P S V A A T L
 2651 GCTCAGGGCT ATAAGGTGCT AGTACTCAAC CCTCTGTGTTG CTGCAACACT
 CGAGTCCCGA TATTCCACGA TCATGAGTTG GGGAGACAAC GACGTTGTGA

+2 G F G A Y M S K A H G I D P N I
 2701 GGGCTTGCT GCTTACATGT CCAAGGCTCA TGGGATCGAT CCTAACATCA
 CCCGAAACCA CGAATGTACA GGTTCCAGT ACCCTAGCTA GGATTGTAGT

+2 R T G V R T I T T G S P I T Y S T
 2751 GGACCGGGGT GAGAACAAATT ACCACTGGCA GCCCCATCAC GTACTCCACC
 CCTGGCCCA CTCTGTAA TGGTGACCGT CGGGTAGTG CATGAGGTGG

+2 Y G K F L A D G G C S G G A Y D I
 2801 TACGGCAAGT TCCCTGCCGA CGGGGGGTGC TCGGGGGCG CTTATGACAT
 ATGCCGTTCA AGGAACGGCT CGCCGCCACG AGCCCCCGC GAATACTGTA

+2 I I C D E C H S T D A T S I L G
 2851 ATAATTTGT GACGAGTGCC ACTCCACCGA TGCCACATCC ATCTTGGCA
 TTATTAACCA CTGTCACGG TGAGGTGCCT ACGGTGTAGG TAGAACCCGT

+2 I G T V L D Q A E T A G A R L V V
 2901 TTGGCACTGT CCTTGACCAA GCAGAGACTG CGGGGGCGAG ACTGGTTGTG
 AACCGTGACA GGAAGTGGT CGTCTGTGAC GCCCCGCTC TGACCAACAC

+2 L A T A T P P G S V T V P H P N I
 2951 CTCGCCACCG CCACCCCTCC GGGCTCCGTC ACTGTGCCCC ATCCCAACAT
 GAGCGGTGCG GGTGGGGAGG CCCGAGGCAG TGACACGGGG TAGGGTTGTA

+2 E E V A L S T T G E I P F Y G K
 3001 CGAGGAGGTT GCTCTGTCCA CCACCGGAGA GATCCCTTT TACGGCAAGG
 GCTCCTCCAA CGAGACAGGT GGTGGCTCT CTAGGGAAAA ATGCCGTCC

+2 A I P L E V I K G G R H L I F C H
 3051 CTATCCCCCT CGAAGTAATC AAGGGGGGGA GACATCTCAT CTTCTGTGAT
 GATAGGGGGGA GCTTCATTAG TTCCCCCTCT CTGTAGAGTA GAAGACAGTA

+2 S K K K C D E L A A K L V A L G I
 3101 TCAAAGAAGA AGTGCACGA ACTCGCCGCA AAGCTGGTCG CATTGGGCAT
 AGTTCTTCT TCACGCTGCT TGAGCGCGT TTCGACCGAG GTAACCCGTA

+2 N A V A Y Y R G L D V S V I P T
 3151 CAATGCCGTG GCCTACTACC GCGGTCTTGA CGTGTCCGTC ATCCCGACCA
 GTTACGGCAC CGGATGATGG CGCCAGAACT GCACAGGCAG TAGGGCTGGT

+2 S G D V V V V A T D A L M T G Y T
 3201 GCGGGATGT TGTCGTGCGT GCAACCGATG CCCTCATGAC CGGCTATAACC
 CGCCGCTACA ACAGCAGCAC CGTTGGCTAC GGGAGTACTG GCCGATATGG

pCMV-NS34A

FIGURE 9 - Page 5

+2 G D F D S V I D C N T C V T Q T V
 3251 GGCAGACTCG ACTCGGTGAT AGACTGCAAT ACGTGTGTC CCCAGACAGT
 CCCCTGAAGC TGAGCCACTA TCTGACGTTA TGACACAGT GGGTCTGTC

+2 D F S L D P T F T I E T I T L P
 3301 CGATTTCAGC CTTGACCTA CCTTCACCAT TGAGACAATC ACGCTCCCCC
 GCTAAAGTCG GAAGTGGAT GGAAGTGGTA ACTCTGTTAG TGCGAGGGG

+2 Q D A V S R T Q R R G R T G R G K
 3351 AAAGATGCTGT CTCCCGCACT CAACGTCGGG GCAGGACTGG CAGGGGGAAAG
 TTCTACGACA GAGGGCGTGA GTTGAGCCCC CGTCCCTGACC GTCCCCCTTC

+2 P G I Y R F V A P G E R P S G M F
 3401 CCAGGCATCT ACAGATTGT GGCACCCGGG GAGGCCCT CGGGCATGTT
 GGTCCGTAGA TGTCTAAACA CCGTGGCCCC CTCGCGGGGA GGCGTACAA

+2 D S S V L C E C Y D A G C A W Y
 3451 CGACTCGTCC GTCCCTGTG AGTGTATGA CGCAGGCTGT GCTTGGTATG
 GCTGAGCAGG CAGGAGACAC TCACGATACT GCGTCCGACA CGAACCATAC

+2 E L T P A E T T V R L R A Y M N T
 3501 AGCTCACGCC CGCCGAGACT ACAGTTAGGC TACGAGCGTA CATGAACACC
 TCGAGTGCAGG GCGGCTCTGA TGTCAATCCG ATGCTCGCAT GTACTTGTGG

+2 P G L P V C Q D H L E F W E G V F
 3551 CCGGGGCTTC CCGTGTGCCA GGACCATCTT GAATTTGGG AGGGCGTCTT
 GGCCCCGAAG GGCACACGGT CCTGGTAGAA CTAAAACCC TCCCAGAGAA

+2 T G L T H I D A H F L S Q T K Q
 StuI

 3601 TACAGGCCTC ACTCATATAG ATGCCCACTT TCTATCCCAG ACAAAGCAGA
 ATGTCCGGAG TGAGTATATC TACGGGTGAA AGATAGGGTC TGTTCGTCT

+2 S G E N L P Y L V A Y Q A T V C A
 3651 GTGGGGAGAA CCTTCCTAC CTGGTAGCGT ACCAAGCCAC CGTGTGCGCT
 CACCCCTCTT GGAAGGAATG GACCATCGCA TGTTTCGGTG GCACACCGCA

+2 R A Q A P P P S W D Q M W K C L I
 3701 AGGCTCAAG CCCCTCCCCC ATCGTGGGAC CAGATGTGGA AGTGTGTTGAT
 TCCCGAGTTC GGGGAGGGGG TAGCACCCCTG GTCTACACCT TCACAAACTA

+2 R L K P T L H G P T P L L Y R L
 3751 TCGCCTCAAG CCCACCTCC ATGGGCCAAC ACCCTGCTA TACAGACTGG
 AGCGGAGTTC GGGTGGGAGG TACCCGGTTG TGGGGACGAT ATGTCTGACC

+2 G A V Q N E I T L T H P V T K Y I
 3801 GCGCTGTTCA GAATGAAATC ACCCTGACGC ACCCAGTCAC CAAATACATC
 CGCGACAAGT CTTACTTTAG TGGGACTGCG TGGGTCACTG GTTATGTAG

+2 M T C M S A D L E V V T S T W V L
 3851 ATGACATGCA TGTCGGCCGA CCTGGAGGTC GTCACGAGCA CCTGGGTGCT
 TACTGTACGT ACAGCCGCT GGACCTCCAG CAGTGCTCGT GGACCCACGA

+2 V G G V L A A L A A Y C L S T G
 3901 CGTTGGCGGC GTCCCTGGCTG CTTGGCCGC GTATTGCGTG TCAACAGGCT
 GCAACCGCCG CAGGACCGAC GAAACCGGGCG CATAACGGAC AGTTGTCCGA

pCMV-NS34A
FIGURE 9 - Page 6

```

+2 C V V I V G R V V L S G K P A I I
3951 CGGTGGTCAT AGTGGGCAGG GTCGTCTTGT CCGGGARGCC GGCAATCATA
     CGCACCAAGTA TCACCCGTCC CAGCAGAACAA GGCCCTTCGG CCGTTAGTAT

+2 P D R E V L Y R E F D E M E E C
4001 CCTGACAGGG AAGTCCTCTA CCGAGAGTTC GATGAGATGG AAGAGTGCTA
     GGACTGTCCC TTCAGGAGAT GGCTCTCAAG CTACTCTACC TTCTCACGAT

BamHI      MluI
----- -----
4051 GGATCCACTA CGCGTTAGAG CTCGCTGATC AGCCTCGACT GTGCCTTCTA
     CCTAGGTGAT GCGCAATCTC GAGCGACTAG TCGGAGCTGA CACGGAAGAT

4101 GTTGCCAGCC ATCTGTTGTT TGCCCCCTCCC CGGTGCCCTTC CTTGACCCCTG
     CAACGGTCGG TAGACAACAA ACGGGGAGGG GGCACGGAAG GAACCTGGAC

4151 GAAGGTGCCA CTCCCACGT CTTTCTAA TAAAATGAGG AAATTGCATC
     CTTCCACGGT GAGGGTGACA GGAAAGGATT ATTTTACTCC TTTAACGTAG

4201 GCATTGTCTG AGTAGGTGTC ATTCTATTCT GGGGGGTGGG GTGGGGCAGG
     CGTAACAGAC TCATCCACAG TAAGATAAGA CCCCCCACCC CACCCCGTCC

4251 ACAGCAAGGG GGAGGATTGG GAAGACAATA GCAGGCATGC TGGGGAGCTC
     TGTGTTCCC CCTCTTAACC CTTCTGTT CGTCCGTACG ACCCTCGAG

4301 TTCCGCTTCC TCGCTCACTG ACTCCGCTGCG CTCGGTCGTT CGGCTGCGC
     AAGGCGAAGG AGCGAGTGCAG TGAGCGACGC GAGCCAGCAA GCCGACGCCG

4351 GAGCGGTATC AGCTCACTCA AAGGCGTAA TACGGTTATC CACAGAATCA
     CTCCCATAG TCGAGTGAGT TTCCGCCATT ATGCCAATAG GTGTCTTAGT

4401 GGGGATAACG CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG
     CCCCTATTGC GTCCCTTCTT GTACACTCGT TTTCCGGTCG TTTTCCGGTC

4451 GAACCGTAA AAGGCGCGT TGCTGGCGTT TTCCCATAGG CTCCGCCCCC
     CTTGGCATT TTCCGGCGCA ACGACCGCAA AAAGGTATCC GAGGCGGGGG

4501 CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG
     GACTGCTCGT AGTGTTTTA GCTGGAGTT CAGTCTCCAC CGCTTTGGGC

4551 ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAAGCT CCCTCGTGC
     TGTCCTGATA TTTCTATGGT CCGCAAAGGG GGACCTTCGA GGGAGCACGC

4601 CTCTCCTGTT CCGACCCCTGC CGCTTACCGG ATACCTGTCC GCCTTCTCC
     GAGAGGACAA GGCTGGAGC GCGAATGGCC TATGGACAGG CGGAAAGAGG

4651 CTTCGGGAAG CGTGGCGCTT TCTCAATGCT CACGCTGTAG GTATCTCAGT
     GAAGCCCTTC GCACCGCGAA AGAGTTACGA GTGCGACATC CATAGAGTCA

4701 TCGGTGAGG TCGTTCGCTC CAAGCTGGGC TGTTGTGCACG AACCCCCCGT
     AGCCACATCC AGCAAGCGAG GTTCGACCCCG ACACACGTGC TTGGGGGGCA

4751 TCAGCCCGAC CGCTGGCGCT TATCCGGTAA CTATCGTCTT GAGTCCAACC
     AGTCGGGCTG GCGACCGCGA ATAGGCCATT GATAGCAGAA CTCAGGTTGG

4801 CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT
     GCCATTCTGT GCTGAATAGC GGTGACCGTC GTCGGTGACC ATTGTCTAA

```

pCMV-NS34A

FIGURE 9 - Page 7

4851 AGCAGAGCGA GGTATGTAAG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC
 TCGTCTCGCT CCATACATCC GCCACGGATGT CTCAGAACT TCACCACCGG
 4901 TAACTACGGC TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA
 ATTGATGCCG ATGTGATCTT CCTGTCATAA ACCATAGACG CGAGACGACT
 4951 AGCCAGTTAC CTTCGGAAAA AGAGTGGTA GCTCTTGATC CGGAAACAA
 TCGGTCAATG GAAGCCTTT TCTCAACCAT CGAGAACTAG GCCGTTGTT
 5001 ACCACCGCTG GTAGCGGTGG TTTTTTGTG TCGAAGCAGC AGATTACCGC
 TGGTGGCGAC CATCGCCACC AAAAAAACAA ACGTTCGTCG TCTAATGCGC
 5051 CAGAAAAAAA GGATCTCAAG AAGATCCTT GATCTTTCT ACGGGGTCTG
 GTCTTTTTT CCTAGAGTTC TTCTAGGAAA CTAGAAAAGA TGCCCCAGAC
 5101 ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTGGT CATGAGATTA
 TGGAGTCAC CCTGCTTTG AGTGAATTG CCTAAACCA GTACTCTAAT
 5151 TCAAAAGGA TCTTCACCTA GATCCTTTA AATTAAAAAT GAAGTTTAA
 AGTTTTTCT AGAAGTGGAT CTAGGAAAAT TTAATTTTA CTTCAAAATT
 5201 ATCAATCTAA AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT
 TAGTTAGATT TCATATATAC TCATTGAAAC CAGACTGTCA ATGGTTACGA
 5251 TAATCAGTG GGCACCTATC TCAGGGATCT GTCTATTTCG TTCACTCCATA
 ATTAGTCACT CCGTGGATAG AGTCGCTAGA CAGATAAACG AAGTAGGTAT
 5301 GTGCCCTGAC TCCCCGTGCT GTAGATAACT ACGATACGGG AGGGCTTACCC
 CAACGGACTG AGGGGCAGCA CATCTATTGA TGCTATGCC CCCCCAATGG
 5351 ATCTGGCCCC AGTGTGCAA TGATACCGCG AGACCCACGC TCACCGGCTC
 TAGACCGGGG TCACGACGTT ACTATGGCGC TCTGGGTGCG AGTGGCCGAG
 5401 CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT
 GTCTAAATAG TCGTTATTG GTCGGTGCGC CTTCCGGCT CGCGTCTCA
 5451 GGTCCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAAATT GTGCCGGGA
 CCAGGACGTT GAAATAGGGC GAGGTAGTC AGATAATTAA CAACGGCCCT
 5501 AGCTAGAGTA AGTAGTTGCG CAGTTAATAG TTGCGCAAC GTTGTGCCA
 TCGATCTCAT TCATCAAGCG GTCAATTATC AAACGCGTTG CAACAACGGT
 5551 TTGCTACAGG CATCGGGTG TCACGCTGCT CGTTGGTAT GGCTTCATC
 AACGATGTCC GTAGCACAC AGTGCAGCA GCAAACCATA CCGAAGTAAG
 5601 AGCTCCGGTT CCCAACGATC AAGGGGAGTT ACATGATCCC CCATGTTGTG
 TCGAGGCCAA GGGTTGCTAG TTCCGCTCAA TGTACTAGGG GGTACAACAC
 5651 CAAAAAAGCG GTTAGCTCCT TCAGGCTCTCC GATCGTTGTC AGAAGTAAGT
 GTTTTTCCG CAATCGAGGA AGCCAGGAGG CTAGCAACAG TCTTCATTCA
 5701 TGGCCGCACT GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT
 ACCGGCGTCA CAATAGTGAG TACCAATACC GTCGTGACGT ATTAAGAGAA
 5751 ACTGTCATGC CATCCGTAAG ATGCTTTCT GTGACTGGTG AGTACTCAAC
 TGACAGTACG GTAGGCATTC TACGAAAAGA CACTGACCAC TCATGAGTTG

pCMV-NS34A**FIGURE 9 - Page 8**

5801 CAAGTCATTC TGAGAATAGT GTATGCGGCG ACCGAGTTGC TCTTGCCCCGG
GTTCAAG ACTCTTATCA CATAACCGC TGGCTAACG AGAACGGGCC

5851 CGTCAATAACG GGATAATACC GCGCCACATA GCAGAACTTT AAAAGTGC
TCAGTATGC CCTATTATGG CGCGGTGTAT CGTCTGAAA TTTTCACGAG

5901 ATCATTGGAA AACGTTCTTC GGGGGCGAAAA CTCTCAAGGA TCTTACCGCT
TAGTAACCTT TTGCAAGAAG CCCCGCTTT GAGAGTTCCT AGAATGGCGA

5951 GTTGAGATCC AGTCGATGT AACCCACTCG TGACACCAAC TGATCTTCAG
CAACTCTAGG TCAAGCTACA TTGGGTGAGC ACGTGGGTTG ACTAGAAGTC

6001 CATCTTTAC TTTCACCAGC GTTCTGGGT GAGCAAAAC AGGAAGGCAA
GTAGAAAATG AAAGTGGTCG CAAAGACCA CTCGTTTG TCCTTCCCGT

6051 AATGCCGCAA AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT
TTACGGCGTT TTTCCCTTA TTCCCGCTGT GCCTTTACAA CTTATGAGTA

6101 ACTCTTCCTT TTTCATATT ATTGAAGCAT TTATCAGGGT TATTGTCTCA
TGAGAAGGAA AAAGTTATAA TAACTTCGTA AATAGTCCCATAAACAGAGT

6151 TGAGCGGATA CATATTTGAA TGTATTTAGA AAAATAAACAA AATAGGGGT
ACTCGCCTAT GTATAAAACTT ACATAAACTT TTTTATTTGT TTATCCCCAA

6201 CCGCGCACAT TTCCCCGAAA AGTGCACCT GACGTCTAAG AAACCATTAT
GGCGCGTGTAA AAGGGGCTTT TCACGGTGGAA CTGCAGATTC TTTGGTAATA

6251 TATCATGACA TTAACCTATA AAAATAGGCG TATCAGGAGG CCCTTCGTC
ATAGTACTGT AATTGGATAT TTTTATCCGC ATAGTGCCTCC GGGAAAGCAG

FIGURE 10

1

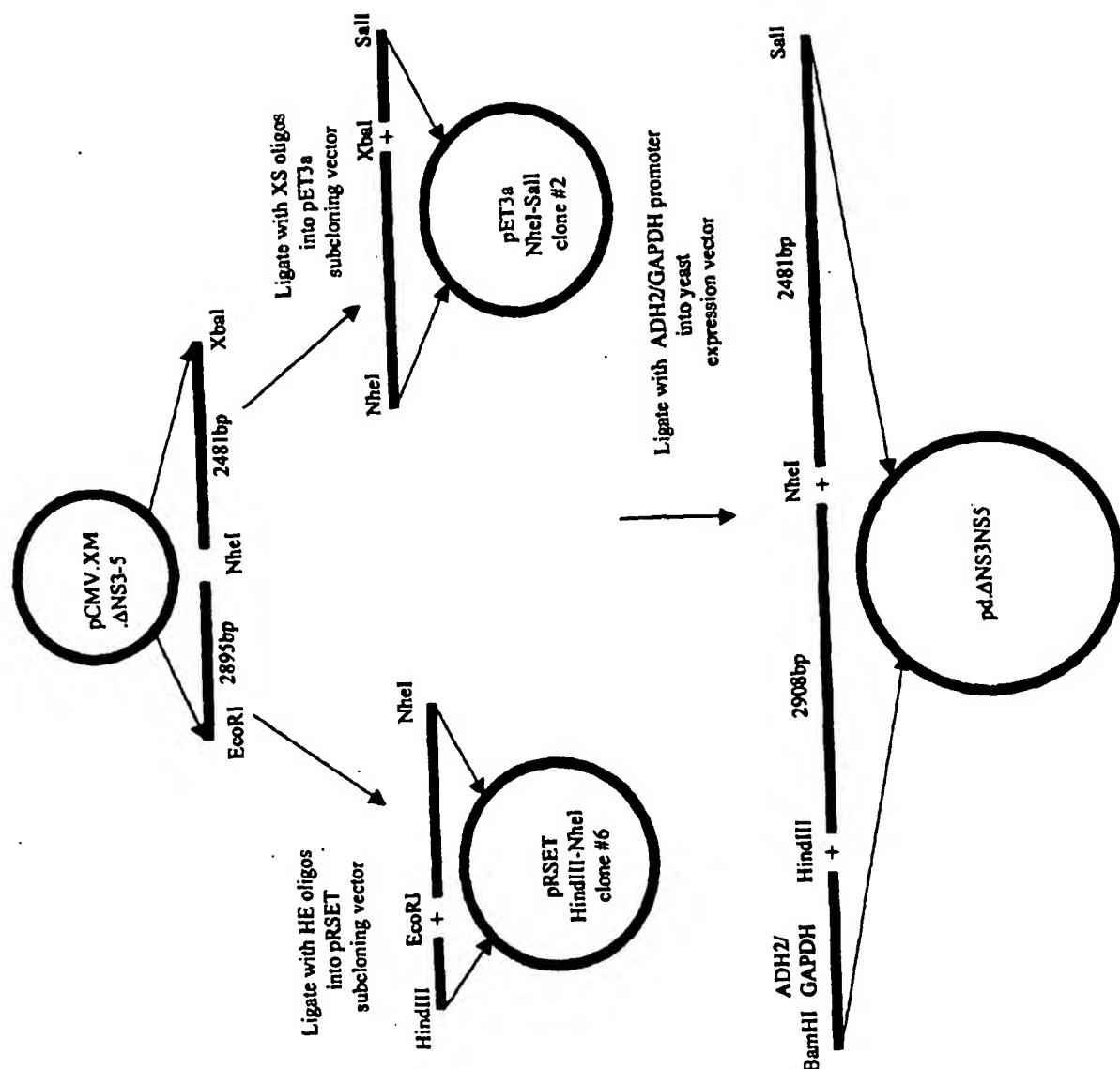


FIGURE 11 - Page 1

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val
 2 AGCTTACAAAACAATTCACCATGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTA
 TCGAATGTTTGTAAAGTGGTACCGACGTATACGTGAGTCCCATAATTCCACGATCAT
 ^ ^ ^
 1 HIND3, 21 NCOI, 30 NDEI, 58 SCAI,

 Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly
 62 CTCACCCCTCTGTGCTGCAACACTGGCTTGGTCTTACATGTCCAAGGCTCATGGG
 GAGTTGGGAGACAACGACGTTGTGACCCAAACCACGAATGTACAGGTTCCGAGTACCC

 Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Gly Ser Pro Ile Thr Tyr
 122 ATCGATCCTAACATCAGGACGGGGTGAGAACAAATTACCACTGGCAGCCCCATACGTAC
 TAGCTAGGATTGTAGTCCTGGCCCCACTCTGTTAATGGTGACCGTCGGGGTAGTGCATG
 ^
 122 CLAI,

 Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile
 182 TCCACCTACGGCAAGTTCCCTGCCGACGGCGGGTGCTCGGGGGCGCTTATGACATAATA
 AGGTGGATGCCGTTCAAGGAACGGCTGCCGCCACGAGCCCCCGGAATACTGTATTAT

 Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu
 242 ATTTGTGACGAGTGCCACTCCACGGATGCCACATCCATCTTGGCATTGGCACTGTCCTT
 TAAACACTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAAACCGTACAGGAA

 Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly
 302 GACCAAGCAGAGACTGCCGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGC
 CTGGTTCGTCTCTGACGCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCG
 ^
 309 ALWN1,

 Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile
 362 TCCGTCACTGTGCCCATCCAAACATCGAGGAGGTTGCTCTGCTCCACCACCGGAGAGATC
 AGGCAGTGACACGGGGTAGGGTTGTAGCTCTCCAACGAGACAGGTGGTGGCCTCTCTAG

 Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe
 422 CCTTTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGAGACATCTCATCTTC
 GGAAAATGCCGTTCCGATAGGGGAGTCATTAGTCCCCCCTCTGTAGAGTAGAAG

 Cys His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn
 482 TGTCAATTCAAAGAAGAAGTGCACGAACTGCCGCAAAGCTGGTGCATTGGCATTCAAT
 ACAGTAAGTTCTTCTCACGCTGCTTGAGCGCGTTCGACCACCGTAACCGTAGTTA

 Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Val
 542 GCCGTGCCCTACTACCGCGGCTTGACGTGCGTCACTCCGACCGAGCGCGATGTTGTC
 CGGCACCGGATGATGGCGCCAGAACTGCACAGGCACTGGCTGGTGCCTACAACAG
 ^ ^
 556 SAC2, 566 DRD1,

 Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp
 602 GTCGTGGCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGAC
 CAGCACCGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTG
 ^
 621 BSPH1,

 Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu

FIGURE 11 - Page 2

662 TGCAATACGTGTGTCACCCAGACAGTCGATTCAGCCTGACCCCTACCTTACCCATTGAG
 ACGTTATGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACGGATGGAAGTGGTAAC
 ^
 722 ThrIleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArg
 ACAATCACGCTCCCCAAGATGCTGTCTCCGCACACTAACGTCGGGCAGGACTGGCAGG
 TGTTAGTGCAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCC
 ^
 782 GlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAsp
 GGGAAAGCCAGGCATCTACAGATTGTGGCACCGGGGAGCGCCCCCTCCGGCATGTTGAC
 CCCTCGGTCCGTAGATGTCTAACACCGTGGCCCCCTCGCGGGGAGGCCGTACAAGCTG
 ^
 822 BGLI, 839 DRD1,
 ^
 SerSerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAla
 842 TCGTCCGTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCC
 AGCAGGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAGTGCAGGGCGG
 ^
 887 SACI,
 ^
 GluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAsp
 902 GAGACTACAGTTAGGCTACGAGCGTACATGAAACACCCGGGCTTCCGTGTGCCAGGAC
 CTCTGATGTCAATCCGATGCTCGCATGTACTTGTGGGCCCGAAGGGCACACGGTCCTG
 ^
 937 SMAI XMAI,
 HisLeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeu
 962 CATCTTGAATTTGGGAGGGCGCTTTACAGGCCTCACTCATATAGATGCCCACTTCTA
 GTAGAACTTAAACCCCTCCCGCAGAAATGTCGGAGTGAGTATATCTACGGGTGAAAGAT
 ^
 991 STUI,
 ^
 SerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrVal
 1022 TCCCAGACAAAGCAGAGTGGGAGAACCTTCTTACCTGGTAGCGTACCAAGCCACCGTG
 AGGGCTGTTCGTCTACCCCTTTGGAGGAATGGACCATCGCATGGTCGGTGGCAC
 ^
 1075 DRA3,
 CysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArg
 1082 TGCCTAGGGCTCAAGCCCTCCCCATCGTGGGACCAGATGTGGAAAGTGGTATTGCG
 ACGCGATCCCGAGTCGGGAGGGTAGCACCTGGTACACCTTCACAAACTAACAGCG
 ^
 LeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsn
 1142 CTCAAGCCCACCCCTCATGGCCAACACCCCTGCTATACAGACTGGCGCTGTTGAGAAT
 GAGTTGGGTGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTA
 ^
 1156 NCOI,
 GluIleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeu
 1202 GAAATCACCCCTGACGCACCCAGTCACCAAATACATCATGACATGGATGCGGCCGACCTG
 CTTTAGTGGGACTGCGTGGGTCACTGGTTATGTAGTACTGTACGTACAGCCGGCTGGAC
 ^ ^ ^ ^
 1236 BSPH1, 1240 DRD1, 1243 AVA3, 1251 EAG1 XMA3, 1256 DRD1,
 ^
 GluValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyr
 1262 GAGGTGCTACGAGCACCTGGGTGCTCGTGGCGGCGCTGGCTGCTTGGCCGCGTAT
 CTCCAGCAGTGCTCGTGGACCCACGAGCAACGCCGAGGACCGACGAAACCGGCGCATA

FIGURE 11 - Page 3

1322 CysLeuSerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAla
 TGCCTGTCAACAGGCTGCGTGGTCATAGTGGCAGGGTCGTCTGTCCGGGAAGCCGGCA
 ACGGACAGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTCGGCCGT
 ^
 1375 NAEI,
 IleIleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGln
 1382 ATCATACCTGACAGGAAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAG
 TAGTATGGACTGTCCCTCAGGAGATGGCTCTAACGCTACTCTACCTCTCACGAGAGTC
 ^
 1391 DRD1,
 HisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeu
 1442 CACTTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTC
 GTGAATGGCATGTAGCTCGTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTCCGGGAG
 GlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsn
 1502 GGCCTCCTGCAGACCGCGTCCCCGTAGGCAGAGGTTATCGCCCCCTGCTGTCCAGACCAAC
 CCGGAGGACGTCTGGCGCAGGGCAGTCCGTCCTAACATAGCAGGGACGACAGGTCTGGTTG
 ^ ^
 1508 PSTI, 1513 TTH3I,
 TrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGln
 1562 TGGCAAAAATCGAGACCTCTGGCGAAGCATATGTGGAACCTTCATCAGTGGGATACAA
 ACCGTTTTGAGCTCTGGAAAGACCCGTTCTGTATAACACCTTGAAGTAGTCACCCATGTT
 ^ ^
 1571 XHOI, 1592 NDEI,
 TyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPhe
 1622 TACTTGGCGGGCTTGTCAACGCTGCCCTGGTAACCCGCCATTGCTTCATTGATGGCTTT
 ATGAACCGCCCGAACAGTTGCGACGGACCATTGGGCGGTAAACGAAGTAACCTACCGAAAA
 ^
 1649 BSTE2,
 ThrAlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGly
 1682 ACAGCTGCTGTCAACGCCACTAACCACTAGCCAAACCTCCTCTCAACATATTGGGG
 TGTCGACGACAGTGGTGGGTGATTGGTATCGGTTGGGAGGAGAAGTTGTATAACCC
 ^
 1683 ALWN1 PVU2,
 GlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGly
 1742 GGGTGGGTGGCTGCCAGCTGCCGCCCCCGTGCCTACTGCCTTGTGGCGCTGGC
 CCCACCCACCGACGGTCGAGCGGCGGGGCCACGGCGATGACGAAACACCCGCGACCG
 ^
 1800 ESP1,
 LeuAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAla
 1802 TTAGCTGGCGCCGCATCGGCAGTGGACTGGGAAGGTCTCATAGACATCCTGCA
 AATCGACCGCGGGCGGTAGCGTCACAACCTGACCCCTCCAGGAGTATCTGTAGGAACGT
 ^
 1808 KAS1 NARI,
 GlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluVal
 1862 GGGTATGGCGCGGGCGTGGCGGGAGCTTGTGGCATTCAAGATCATGAGCGGTGAGGTC
 CCCATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAG
 ^ ^

FIGURE 11 - Page 4

1884 SACI, 1905 BSPH1,

1922 ProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuVal
 CCCTCCACGGAGGACCTGGTCAATCTACTGCCCGCATCCTCTGCCCGGAGCCCTCGTA
 GGGAGGTGCCTCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCAGGGCCTCGGGAGCAT

1934 TTH3I,

1982 ValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaVal
 GTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTGGCCGGCGAGGGGGCAGTG
 CAGCCGCACCAGACAGTCGTTATGACGCCGTGCAACCGGCCCCCTCCCCGTACAC

2010 NAEI, 2023 SMAI XMAI,

2042 GlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHis
 CAGTGGATGAACCGGCTGATAGCCTCGCCCTCCGGGGAACATGTTCCCCCACGCAC
 GTCACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGGGTGCCTG

2073 SMAI XMAI, 2099 DRA3,

2102 TyrValProGluSerAspAlaAlaArgValThrAlaIleLeuSerSerLeuThrVal
 TACGTGCCGGAGAGCGATGCAGCTGCCCGCTACTGCCATACTCAGCAGCCTCACTGTA
 ATGCACGCCCTCGCTACGTGACGGCGCAGTGACGGTATGAGTCGTCGGAGTGACAT

2121 PVU2,

2162 ThrGlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSer
 ACCCAGCTCCTGAGGCGACTGCACCACTGGATAAGCTGGAGTGTACCACTCCATGCTCC
 TGGGTCGAGGACTCCGCTGACGTGGTCACCTATTGAGCCTCACATGGTGAGGTACGAGG

2165 ALWN1, 2170 MST2,

2222 GlySerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThr
 GGTTCTGGCTAAGGGACATCTGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACC
 CCAAGGACCGATTCCCTGTAGACCTTGACCTATACGCTCCACAACCTCGCTGAAATTCTGG

2226 ECON1,

2282 TrpLeuLysAlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArg
 TGGCTAAAGCTAACGCTCATGCCACAGCTGCCCTGGATCCCTTGTGTCCTGCCAGCGC
 ACCGATTTCGATTGAGTACGGTGTGACGGACCCCTAGGGGAAACACAGGACGGTCGCG

2291 ESP1, 2306 PVU2, 2316 BAMHI,

2342 GlyTyrLysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAla
 GGGTATAAGGGGGTCTGGCGAGGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCT
 CCCATATTCCCCAGACCGCTCCCTGCCGTAGTACGTGAGCGACGGTACACCTCGA

2402 GluIleThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArg
 GAGATCACTGGACATGTCAAAACGGGACGATGAGGATCGTCGGCTACTGGACCTGCAGG
 CTCTAGTGACCTGTACAGTTTGGCTACTCCTAGCAGCCAGGATCCTGGACGTCC

2431 BSAB1, 2447 AVR2, 2454 SSE83871, 2455 PSTI,

2462 AsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrGlyProCysThrProLeu
 AACATGTGGAGTGGGACCTCCCCATTAATGCCTACACCACGGGCCCCGTACCCCCCTT
 TTGTACACCTCACCCCTGGAGGGTAATTACGGATGTGGTCCCCGGGACATGGGGGAA

FIGURE 11 - Page 5

2486 ASE1, 2503 APAI,

2522 ProAlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIle
 CCTGCGCCGAACACTACAGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATA
 GGACGCGGCTTGATGTGCAAGCGCATACTCCCACAGACGTCTCCTATGCACCTCTAT

2559 PSTI,

2582 ArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysPro
 AGGCAGGTGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTAAATGCCCG
 TCCGTCCACCCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTACGGGC

2600 DRA3,

2642 CysGlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPhe
 TGCCAGGTCCCATGCCCGAATTTCACAGAATTGGACGGGTGCCCTACATAGGTTT
 ACGGTCCAGGGTAGCGGGCTAAAAAGTGTCTAACCTGCCAACGCGGATGTATCCAAA

2702 AlaProProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGlu
 GCGCCCCCTGCAAGCCCTGCTGCCGGAGGAGGTATCATCAGAGTAGGACTCCACGAA
 CGCGGGGGGACGTTGGAACGACGCCCTCCATAGTAAGTCTCATCCTGAGGTGCTT

2762 TyrProValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSer
 TACCCGGTAGGGTCGCAATTACCTTGCAGGCCGAACCGGACGTGGCGTGTGACGTCC
 ATGGGCCATCCCAGCGTTAATGAAACGCTCGGGCTTGGCTGACCGCACAACTGCAGG

2763 HGIE2, 2815 AAT2,

2822 MetLeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGly
 ATGCTCACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGCGAAGGTTGGCGAGGGGA
 TAGCAGTGACTAGGGAGGGTATATTGTCGTCTCGCCGGCCGCTTCCAACCGCTCCCCT

2856 EAG1 XMA3,

2882 SerProProSerValAlaSerSerAlaSerGlnLeuSerAlaProSerLeuLysAla
 TCACCCCCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCAAGGCA
 AGTGGGGGGAGACACCGGTGAGGAGCCATCGGTGATAGGCGAGGTAGAGAGTCCGT

2895 BALI, 2909 NHEI,

2942 ThrCysThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrp
 ACTTGCACCGCTAACCATGACTCCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGG
 TGAACGTGGCGATTGGTACTGAGGGACTACGACTCGAGTATCTCGGTTGGAGGATACC

2972 ESP1, 2975 SAC1,

3002 ArgGlnGluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeu
 AGGCAGGAGATGGCGGCAACATCACCAAGGGTGAGTCAGAAAACAAAGTGGTGATTCTG
 TCCGTCCCTCACCGCCGTTGAGTGGCTTCAACTCAGTCTTTGTTCAACCAGAC

3062 AspSerPheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGlu
 GACTCCTTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAA
 CTGAGGAAGCTAGGCGAACACCGCCCTCCTGCTGCCCTCTAGAGGCATGGCGTCTT

3102 BGL2,

FIGURE 11 - Page 6

3122 IleLeuArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyr
 ATCCTGCCGAAGTCTCGGAGATTCGCCAGGCCCTGCCGTTGGCGCGCGACTAT
 TAGGACGCCCTCAGAGCCTCTAACGGGACGGGAAACCCGCGCCGCTGATA
 3149 ALWN1, 3170 EAG1 XMA3,
 AsnProProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGly
 AACCCCCGCTAGTGGAGACGTGGAAAAAGCCCCTACGAACCACCTGTGGTCCATGGC
 TTGGGGCGATCACCTCTGCACCTTTCGGGCTGATGCTGGTGACACCAGGTACCG
 3223 HGIE2, 3235 NCOI,
 CysProLeuProProLysSerProProValProProProArgLysLysArgThrVal
 TGCCCGCTTCCACCTCCAAAGTCCCCTCTGTGCCTCCGCCCTCGGAAGAAGCGGACGGTG
 ACGGGCGAAGGTGGAGGTTTCAGGGGAGGACACGGAGCGGAGCCTCTCGCCTGCCAC
 ValLeuThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGly
 GTCCTCACTGAATCAACCTATCTACTGCCTGGCCGAGCTGCCACCAGAACGCTTGCG
 CAGGAGTGACTTAGTTGGATAGATGACGGAACCGGCTCGAGCGGTGGCTTCGAAACCG
 3338 SACI, 3352 HIND3,
 SerSerSerThrSerGlyIleThrGlyAspAsnThrThrSerSerGluProAlaPro
 AGCTCCTCAACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCCTGGC
 TCGAGGAGTTGAAGGCCGTAATGCCGCTGTTATGCTGTTAGGAGACTCGGGCGGGGA
 SerGlyCysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGly
 TCTGGCTGCCCGACTCCGACGCTGAGTCCTATTCCCATGCCCGCTGGAGGGG
 AGACCGACGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGACCTCCCC
 3443 EAM11051,
 GluProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsn
 GAGCCTGGGATCCGATCTAGCGACGGTCATGGTCAACGGTCAGTAGTGAGGCCAAC
 CTCGGACCCCTAGGCTAGAATCGCTGCCAGTACCGTTGCCAGTCATCACTCCGGTTG
 3490 BAMHI, 3491 BSAB1, 3493 BSPE1,
 AlaGluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrPro
 GCGGAGGATGTCGTGCTGCTCAATGTCTACTCTGGACAGGCCACTCGTCACCCCG
 CGCCTCTACAGCACACGAGTTACAGAACGAGTGTCCCGTGAGCAGTGGGG
 3595 DRA3,
 CysAlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHis
 TGCGCCCGGAAAGAACAGAAACTGCCATCAATGCACAAAGCAACTCGTTGCTACGTAC
 ACCGGCGCCTTCTGTCTTGTGAGGGTAGTTACGTGATTGAGCAACGATGCAAGTG
 3606 SAC2, 3617 ALWN1, 3661 PFLM1,
 HisAsnLeuValTyrSerThrSerArgSerAlaCysGlnArgGlnLysLysValThr
 CACAATTGGTGTATTCCACCACTCACGAGTGCTGCCAAAGGCAGAACGAAAGTCACA
 GTGTTAAACCACATAAGGTGGTGGAGTGCCTCACGAACGGTTCCGTCTTCAGTGT
 3687 DRA3,
 PheAspArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAla

FIGURE 11 - Page 7

3722 TTTGACAGACTGCAAGTTCTGGACAGCATTACCAGGACGTACTCAAGGAGGTTAAAGCA
AAACTGTCTGACGTCAAGACCTGCGGTATGGTCCTGCATGAGTTCTCCAAATTCGT
AlaAlaSerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrPro
3782 GCGGCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCC
CGCCGCAGTTTCACTCCGATGAACTAGGCATCTCCTCGAACGTCGGACTGCGGG
3822 HIND3,
ProHisSerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArg
3842 CCACACTCAGCAAATCCAAGTTGGTTATGGGGAAAAGACGTCCGTTGCCATGCCAGA
GGTGTGAGTCGGTTAGGTCAAACCAATAACCCGTTCTGCAGGCAACGGTACGGTCT
3881-AAT2, 3896 BGLI,
LysAlaValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrPro
3902 AAGGCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGAAAGACAATGTAACACCA
TTCCGGCATTGGGTGTAGTGAGGCACACCTTCTGAAAGACCTTCTGTTACATTGTGGT
IleAspThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGly
3962 ATAGACACTACCACATGGCTAACGAGGTTTCTGCTTCAGCCTGAGAAGGGGGT
TATCTGTGATGGTAGTACCGATTCTGCTCAAAGACGCAAGTCGGACTCTCCCCCA
ArgLysProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMet
4022 CGTAAGCCAGCTCGTCTCATCGTGTCCCCTGCGCTGCGGTGCGAAAAGATG
GCATTGGTCGAGCAGAGTAGCACAAAGGGCTAGACCCGCACGCGCACACGCTTTCTAC
AlaLeuTyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPhe
4082 GCTTTGTACGACGTGGTTACAAAGCTCCCTTGCCGTGATGGGAAGCTCCTACGGATTG
CGAAACATGCTGCACCAATGTTGAGGGGAAACCGGCACTACCCTCGAGGATGCCAAAG
GlnTyrSerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThr
4142 CAATACTCACCAGGACAGCGGGTTGAATTCTCGTCAAGCGTGGAAAGTCCAAGAAAACC
GTTATGAGTGGTCCTGTCGCCAACTTAAGGAGCACGTTGACCTTCAGGTTCTTTGG
4166 ECORI,
ProMetGlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIle
4202 CCAATGGGTTCTCGTATGATACCCGCTGCTTGTACTCCACAGTCAGTGAGAGCGACATC
GGTTACCCCAAGAGCATACTATGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAG
4235 DRD1, 4242 ALWN1,
ArgThrGluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIle
4262 CGTACGGAGGAGGCAATCTACCAATGTTGACCTCGACCCCCAAGCCCGTGGCCATC
GCATGCCCTCCCGTTAGATGGTTACAACACTGGAGCTGGGGTTGGCGCACCGGTAG
4307 BGLI, 4314 BALI,
LysSerLeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsn
4322 AAGTCCCTCACCGAGAGGCTTATGTTGGGGCCCTCTTACCAATTCAAGGGGGAGAAC
TTCAGGGAGTGGCTCTCGAAATACAACCCCCGGGAGAATGGTTAAGTCCCCCTTTG
4351 APAI,
CysGlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeu
4382 TGGCGCTATCGCAGGTGCCGCCGAGCGGGTACTGACAACTAGCTGGTAACACCCCTC

FIGURE 11 - Page 8

ACGCCGATCGTCCACGGCGCCTGCCGCATGACTGTTGATGACACCATTGTGGAG
 4442 ThrCysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMet
 ACTTGCTACATCAAGGCCGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCAG
 TGAACGATGTAGTCCGGGCCGTCGGACAGCTCGGCTCCGAGGTCCCTGACGTGGTAC
 4458 SMA1 XMA1,
 4502 LeuValCysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAla
 CTCGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGTCCAGGAGGACGCG
 GAGCACACACCGCTGCTGAATCAGCAATAGACACTTCGCGCCCCCAGGTCCCTGCGC
 4514 DRD1, 4517 TTH3I,
 4562 AlaSerLeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspPro
 GCGAGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCTGGGGACCCC
 CGCTCGGACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGG
 4622 ProGlnProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAla
 CCACAACCAAGAACATCGACTTGGAGGCTCATAACATCATGCTCCTCAAACGTGTAGTCGCC
 GGTGTTGGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCG
 4643 SAC1,
 4682 HisAspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAla
 CACGACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCG
 GTGCTGCCCGCACCTTCTCCAGATGATGGAGTGGGACTGGGATGTTGGGGGAGCGC
 4737 NRUI,
 4742 ArgAlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIle
 AGAGCTGCGTGGGAGACAGCAAGACACACTCCAGTCATTCTGGCTAGGCAACATAATC
 TCTCGACGCCACCCCTCTGCTGTTCTGTGAGGTCACTTAAGGACCGATCCGTTATTAG
 4802 MetPheAlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeu
 ATGTTGGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTCTTAGCGTCCTT
 TACAAACGGGGGTGTGACACCCGCTCTACTATGACTACTGGTAAAGAAATCGCAGGAA
 4812 PFLM1, 4813 DRA3,
 4862 IleAlaArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSer
 ATAGCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCC
 TATCGGTCCTGGTCGAACCTGTCCGGAGCTAACGCTCTAGATGCCCGGACGATGAGG
 4899 BGL2,
 4922 IleGluProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSer
 ATAGAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTC
 TATCTTGGTGAACCTAGATGGAGGTTAGTAAGTTCTGAGGTACCGGAGTCGCGTAAAAGT
 4960 NCOI,
 4982 LeuHisSerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGly
 CTCCACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCCTCAGAAAACCTGGG
 GAGGTGTCATGAGAGGTCCACTTAGTTATCCCACCGGGTACGGAGTCGGTAAACCC
 5021 SPHI, 5041 KPN1,

FIGURE 11 - Page 9

5042 ValProProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAla
 GTACCGCCCTTGCAGCTTGGAGACACCGGGCCGGAGCGTCCCGCCTAGGCTCTGGCC
 CRTGGCGGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCGCGATCCGAAGACCGG
 ^
 5070 APAI, 5097 BALI,
 ArgGlyGlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLys
 5102 AGAGGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTCAACTGGCAGTAAGAACAAAG
 TCTCCTCCGTCCCACGGTATAACCGTTATGGAGAAGTTGACCCGTATTCTTGTTC
 ^
 5119 NDEI,
 LeuLysLeuThrProIleAlaAlaAlaGlyGlnLeuAspLeuSerGlyTrpPheThrAla
 5162 CTCAAACTCACTCCAATAGCGGCCGCTGGCAGCTGGACTTGTCCGCTGGTCACGGCT
 GAGTTTGAGTGAGGTTATCGCCGGCGACGGTCACTGAACAGGCCGACCAAGTGCCGA
 ^ ^ ^
 5180 NOTI, 5181 EAG1 XMA3, 5188 BALI, 5192 PVU2,
 GlyTyrSerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrp
 5222 GGCTACAGCGGGGAGACATTATCACAGCGTGTCTCATGCCCGCCCCGCTGGATCTGG
 CCGATGTCGCCCCCTGTAAAATAGTGTGGCACAGAGTACGGGCCGGCGACCTAGACC
 ^
 5246 DRA3,
 PheCysLeuLeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgOP
 5282 TTTGCCTACTCCTGCTTGCAGGGGTAGGCATCACCTCCTCCCCACCGATGAAGG
 AAAACGGATGAGGACGAACGACGTCCCCATCCGTAGATGGAGGAGGGTTGGCTACTTCC
 ^
 5301 PSTI, 5331 HGIE2,
 5342 TTGGGGTAAACACTCCGGCCTAAAAAAAAAAAAAATCTAGAACCCGAGTCGAC
 AACCCCATTTGTGAGGCCGGATTTTTTTTTAGATCTGGGCTCAGCTG
 ^ ^
 5378 XBAI, 5390 SALI,

FIGURE 12

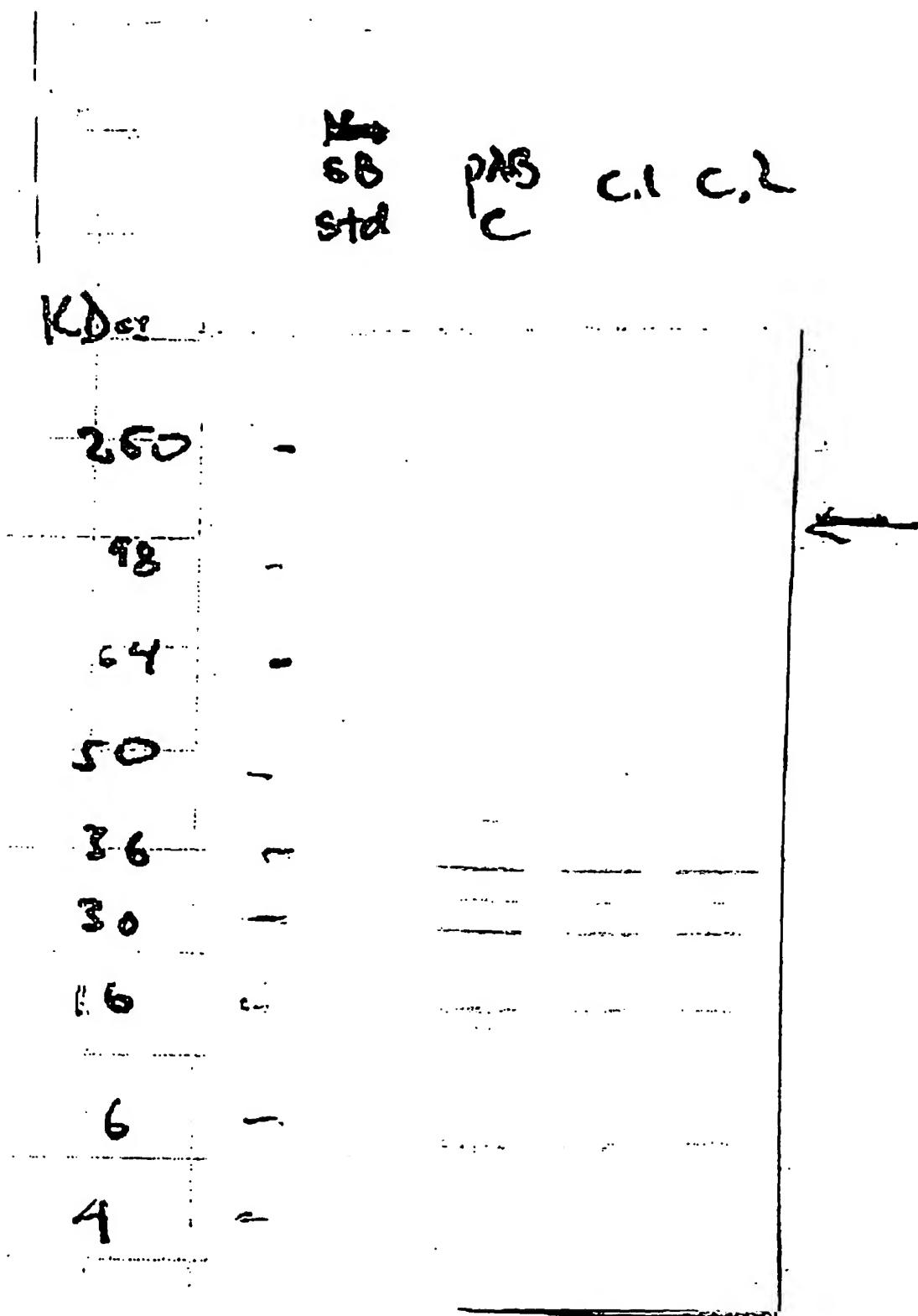


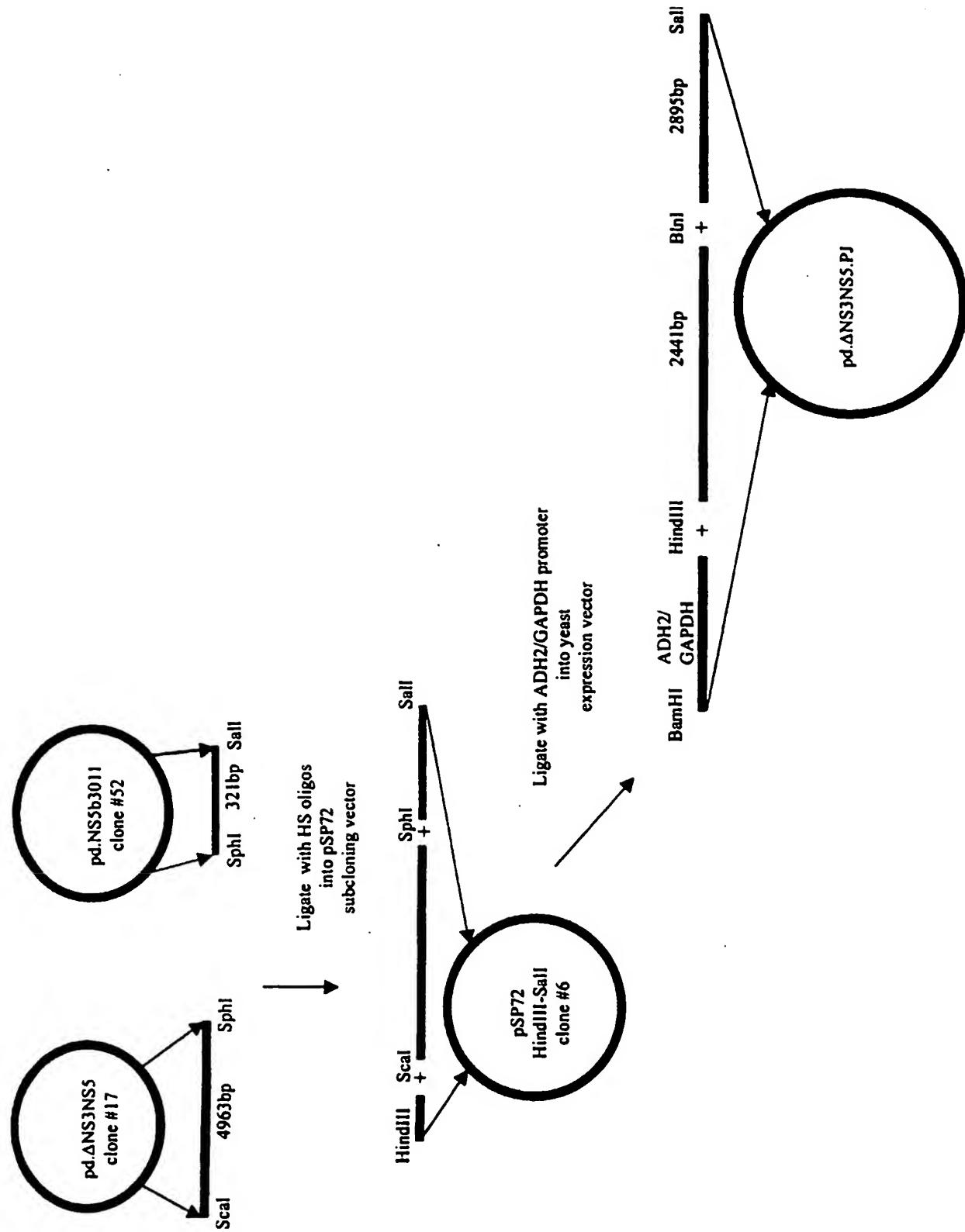
FIGURE 13

FIGURE 14 - Page 1

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn
 2 AGCTTACAAAACAAAATGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAAC
 TCGAATGTTGTTTACCGACGTACGTCGAGTCCCAGATTCCACGATCATGAGTTG
 ^ ^ ^
 1 HIND3, 24 NDEI, 52 SCAI,

 Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp
 62 CCCTCTGTTGCTGCAACACTGGGCTTGGTGCCTACATGTCCAAGGCTCATGGGATCGAT
 GGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTA
 ^
 116 CLAI,

 Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Gly Ser Pro Ile Thr Tyr Ser Thr
 122 CCTAACATCAGGACCGGGGTGAGAACATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGG

 Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys
 182 TACGGCAAGTCCCTGCCGACGGCGGGTGCCTCGGGGGCGCTTATGACATAATAATTGTT
 ATGCCGTTCAAGGAACGGCTGCCGCCACGAGCCCCCGCGAATACTGTATTATTAAACA

 Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln
 242 GACGAGTGCACACTCCACGGATGCCACATCCATCTGGGATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTAGGTAGAACCGTAACCGTGACAGGAACGGTT

 Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val
 302 GCAGAGACTGCAGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGAGGCCGAGGCAG
 ^
 303 ALWN1,

 Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe
 362 ACTGTGCCCATCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTT
 TGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTAGGGAAAA

 Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His
 422 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTGTGTCAT
 ATGCCGTTCCGATAGGGGAGCTTCATTAGTCCCCCTCTGTAGAGTAGAACAGACTA

 Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val
 482 TCAAAGAAGTGCAGGAACCTGCCGAAAGCTGGTGCCTGGCATCAATGCCGTG
 AGTTTCTTCTCACGCTGCTTGAGCGCGTTCGACCAGCGTAACCGTAGTTACGGCAC

 Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Val Val Val
 542 GCCTACTACCGCGGTCTGACGTGTCGTCATCCGACCAGCGCGATGGTGTGCGTGTG
 CGGATGATGGCGCCAGAACTGCAACAGCAGTAGGGCTGGTGCCTGCTACAACAGCAGCAC
 ^ ^
 550 SAC2, 560 DRD1,

 Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn
 602 GCAACCGATGCCCTCATGACCGGCTATACCGGCAGCTCGACTCGGTGATAGACTGCAAT
 CGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACGTTA
 ^
 615 BSPH1.

FIGURE 14 - Page 2

662 ACGTGTGTCACCCAGACAGTCGATTCAGCCTGACCCTACCTTCACCATTTGAGACAATC
 TGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCGTAG
 ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 722 ACGCTCCCCCAAGATGCTCTCCCGACTCAACGTCGGGGCAGGACTGGCAGGGGGAAAG
 TCGGAGGGGGTCTACGACAGAGGGCGTGAGTTGCAGCCCCGCTGACCGTCCCCCTTC
 ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 762 CCAGGCATCTACAGATTGTGGCACGGGGGAGGCCCTCCGGCATGTTGACTCGTCC
 GGTCCTGAGATGTCTAACACCGTGGCCCCCTCGGGGGAGGCCGTACAAGCTGAGCAGG
 816 BGLI, 833 DRD1,
 ValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThr
 842 GTCCCTGTGAGTGTCTATGACGCAGGCTGTGCTGGTATGAGCTACGCCCGAGACT
 CAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAGTGCAGGGCGCTCTGA
 881 SACI,
 ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 902 ACAGTTAGGCTACGAGCGTACATGAACACCCGGGCTTCCCGTGTGCCAGGACCATCTT
 TGTCATCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCTGGTAGAA
 931 SMAI XMAI,
 GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln
 962 GAATTTGGGAGGGCGTCTTACAGGCCTCACTCATATAGATGCCACTTCTATCCCAG
 CTAAAAACCCCTCCCGCAGAAATGTCGGAGTGACTATATCTACGGGTGAAAGATAGGGTC
 985 STUI,
 ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
 1022 ACAAAAGCAGAGTGGGAGAACCTTCCTACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
 TGTTCGTCTACCCCTCTGGAGGAATGGACCATCGCATGGTTCGGTGGCACACCGCGA
 1069 DRA3,
 ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLys
 1082 AGGGCTCAAGCCCCCTCCCCATCGTGGGACCAGATGTGGAAGTGTGTTGATTCGCCTCAAG
 TCCCAGAGTTCGGGAGGGGGTAGCACCTGGTCTACACCTTCACAAACTAACGGAGTTC
 ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 1142 CCCACCCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCCTGTTGAGAATGAAATC
 GGGTGGGAGGTACCGGTTGTGGGACGATATGTCTGACCCCGACAAGTCTTACTTTAG
 1150 NCOI,
 ThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluVal
 1202 ACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCGACCTGGAGGTC
 TGGGACTGCGTGGGTCACTGGTTATGTAGTACTGTACGTACAGCCGGTGGACCTCCAG
 1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,
 ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 1262 GTCACGAGCACCTGGGTGCTCGTTGGCGCGCTGGCTGCTTGGCCGCGTATTGCCCTG
 CAGTGCTCGTGGACCCACGAGCAACCGCCGCAAGGACCGACGAAACCGGCGATAACGGAC

FIGURE 14 - Page 3

1322 SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIle
 TCAACAGGCTGCGTGGTCATAGTGGCAGGGCGTCTGTCCGGGAAGCCGGCAATCATA
 AGTTGTCCGACGCCAGTATCACCGTCCCAGCAGAACAGGCCCTCGGCCGTTAGTAT
 ^
 1369 NAEI,
 ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
 1382 CCTGACAGGGAAAGTCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTA
 GGACTGTCCCTTCAGGAGATGGCTCTAAGCTACTCTACCTCTCACGAGAGTCGTGAAT
 ^
 1385 DRD1,
 ProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAACGCCCTGGCCTC
 GGCATGTAGCTCGTCCCTACTACGAGCGGCTCGTCAAGTTCGTTCCGGGAGCCGGAG
 LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCGTCAGTGGCAGACCAACTGGCAA
 GACGCTGGCGCAGGGCAGTCCGTCCTCAATAGCGGGGACGACAGGTCTGGTTGACC GTT
 ^
 1502 PSTI, 1507 TTH3I,
 LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
 1562 AAACTCGAGACCTTCTGGCGAAGCATAATGTGAACTTCATCATAGTGGATACAATACTTG
 TTTGAGCTCTGAAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATGAAC
 ^
 1565 XHOI, 1586 NDEI,
 AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla
 1622 GCGGGCTTGTCAACGCTGCCGGTAACCCGCCATTGCTTCATGATGGCTTTACAGCT
 CGCCCGAACAGTTCGCGACGGACCATTGGGGCGTAACGAAAGTAACCGAAAATGTCGA
 ^
 1643 BSTE2, 1677 ALWN1 PVU2,
 AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp
 1682 GCTGTCACCAGCCCCTAAACCACTAGCCAAACCCCTCTTCACATATTGGGGGGGGTGG
 CGACAGTGGTCGGGTGATTGGTATCGGTTGGGAGGAGATTGTATAACCCCCCCCACC
 ValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAla
 1742 GTGGCTGCCAGCTGCCGGGGGGTGCCGCTACTGCCTTGTGGCGCTGGCTTAGCT
 CACCGACGGGTGAGCGGGGGGCCACGGCGATGACGGAAACACCCGCGACCGAACATCGA
 ^
 1794 ESP1,
 GlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyr
 1802 GGCGCCGCCATCGCAGTGTGGACTGGGAAGGTCTCATAGACATCCTGCAGGGTAT
 CCGCGCGGTAGCGTCACAAACCTGACCCCTCCAGGAGTATCTGTAGGAACGTCCCATA
 ^
 1802 KAS1 NARI,
 GlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSer
 1862 GGC CGGGCGTGGCGGGAGCTCTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCC
 CCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTAGTACTCGCCACTCCAGGGGAGG
 ^
 1878 SAC1, 1899 BSPH1,

FIGURE 14 - Page 4

1922 ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly
 ACGGAGGACCTGGTCAATCTACTGCCGCATCCTCTGCCGGAGCCCTCCTAGTCGGC
 TGCCTCCTGGACCAGTTAGATGACGGCGTAGGAGAGCGGGCCTCGGAGCATCAGCCG
 ^
 1928 TTH3I,
 ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
 1982 GTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGCCCGGGCAGGGGGCAGTGCAGTGG
 CACCAGACACGTCGTTATGACGCCGGTGCACCGGGCCGCTCCCCGTACGTCACC
 ^ ^
 2004 NAEI, 2017 SMAI XMAI,
 MetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
 2042 ATGAACCGGCTGATAGCCTCGCCTCCGGGGAAACCATGTTCCCCCACGCACACTACGTG
 TACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGGGTGCCTGATGCAC
 ^
 2067 SMAI XMAI, 2093 DRA3,
 ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
 2102 CCGGAGAGCGATGCACTGCCCGCTACTGCCATACTCAGCAGCCTCACTGTAACCCAG
 GCCCTCTCGCTACGTCACGGCGCAGTACGGTATGAGTCGTCGGAGTGACATTGGGTC
 ^
 2115 PVU2, 2159 ALWN1,
 LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
 2162 CTCTGAGGCGACTGCACCACTGGATAAGCTGGAGTGTACCACTCCATGCTCCGGTTCC
 GAGGACTCCGCTGACGTGGCACCTATTGAGCCTCACATGGTGAGGTACGAGGCAAGG
 ^
 2164 MST2, 2220 ECON1,
 TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
 2222 TGGCTAAGGGACATCTGGACTGGATATGCGAGGTGTGAGCGACTTTAACGACTGGCTA
 ACCGATTCCCTGTAGACCTACGCTCCACAACACTCGCTGAAATTCTGGACCGAT
 ^
 LysAlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyr
 2282 AAAGCTAACGCTCATGCCACAGCTGCCCTGGATCCCCTTGTGCTGCCAGCGCGGGTAT
 TTTCGATTCGAGTACGGTTCGACGGACCCCTAGGGAAACACAGGACGGTCGCCGCCCATA
 ^
 2285 ESP1, 2300 PVU2, 2310 BAMHI,
 LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 2342 AAGGGGTCTGGCAGGGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATC
 TTCCCCCAGACCGCTCCCTGCCGTAGTACGTGTAGCGACGGTACACCTCGACTCTAG
 ^
 ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 2402 ACTGGACATGTCAAAACGGGACGATGAGGATCGCTGGCTTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTGTAC
 ^ ^ ^
 2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,
 TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla
 2462 TGGAGTGGGACCTCCCCATTAATGCCACACCACGGGCCCCCTGTACCCCCCTCTGCG
 ACCTCACCCCTGGAAGGGTAATTACGGATGTGGTGCCCGGGACATGGGGGAAGGACGC
 ^
 2480 ASE1, 2497 APA1,
 ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln

FIGURE 14 - Page 5

2522 CCGAACTACACGTTCGCGTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCGAG
GGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTATGCACCTCTATTCCGTC

2553 PSTI,
ValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGln

2582 GTGGGGACTTCACTACGTGACGGGTATGACTACTGACAATCTAAATGCCGTGCCAG
CACCCCCCTGAAGGTATGCACTGCCCATACTGATGACTGTTAGAATTACGGGCACGGTC

2594 DRA3,
ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro

2642 GTCCCATCGCCCCGAATTTTCACAGAATTGGACGGGTGCGCCTACATAGGTTGCC
CAGGTAGCGGGCTTAAAAAGTGTCTAACCTGCCAACGCGGATGTATCCAAACGCGGG

2702 ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
CCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTAGAGTAGGACTCCACGAATACCG
GGGACGTTGGAACGACGCCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGC

2757 HGIE2,
ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu

2762 GTAGGGTCGCAATTACCTTGCAGGCCGAACCGGACGTGGCGTGTGACGTCCATGCTC
CATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCAGCACAACTGCAGGTACGAG

2809 AAT2,
ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro

2922 ACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGCGAAGGTTGGCGAGGGGATCACCC
TGAAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTCCAAACGCTCCCTAGTGGG

2950 EAG1 XMA3,
ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys

2982 CCCTGTGGCCAGCTCTCGGCTAGCCAGCTATCCGCTCCATCTCAAGGCAACTTGC
GGGAGACACCGGTCGAGGAGCCATCGGTCGATAGGCAGGGTAGAGAGTTCCGTTGAACG

2889 BALI, 2903 NHEI,
ThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGln

2942 ACCGCTAACCATGACTCCCTGATGCTGAGCTATAGAGGCCAACCTCTATGGAGGCAG
TGGCGATTGGTACTGAGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCCGTC

2966 ESP1, 2969 SACI,
GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer

3002 GAGATGGCGGCCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTATTCTGGACTCC
CTCTACCCGCCGGTGTAGTGGTCCCAACTCAGTCTTGTGTTACCAACTAACGACTGAGG

3062 PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
TTCGATCCGCTTGTGGCGGAGGAGGACGAGCAGGGAGATCTCCGTACCCGCAGAAATCCTG
AAGCTAGGCGAACACCGCCTCCCTGCTCGCCCTAGAGGCATGGCGTCTTAGGAC

3096 BGL2,
ArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro

3122 CGGAAGTCTCGGAGATTGCCAGGCCCTGCCGTTGGCGCCGGACTATAACCC

FIGURE 14 - Page 6

GCCTTCAGAGCCTCTAAGCGGGTCCGGGACGGCAAACCCGCGCCGGCCTGATATTGGGG
 ~ ~ ~
 3143 ALWN1, 3164 EAG1 XMA3,

· ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGCCCG
 GGCATCACCTCTGCACCTTTCTGGGCTGATGTTGGACACCAGGTACCGACGGGC
 ^ ^
 3217 HGIE2, 3229 NCOI,
 · LeuProProProLysSerProProValProProProArgLysLysArgThrValValLeu
 3242 CTTCACCTCCAAAGTCCCCCTCTGTGCCCTCGGAAGAACGGACGGTGGCTCC
 GAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTCGCCTGCCACCAGGAG
 · ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer
 3302 ACTGAATCAACCTATCTACTGCCTGCCGAGCTGCCACCAAGCTTGGCAGCTCC
 TGACTTAGTTGGATAGATGACGGAACCGGCTCGAGCGGTGGCTTCGAAACCGTCGAGG
 ^ ^
 3332 SACI, 3346 HIND3,
 · SerThrSerGlyIleThrGlyAspAsnThrThrSerSerGluProAlaProSerGly
 3362 TCAACTCCGGCATTACGGCGACAATACGACAACATCCTCTGAGCCCGCCCTCTGGC
 AGTTGAAGGCCGTAATGCCGCTGTTATGCTGTTAGGAGACTCGGGCGGGGAAGACCG
 · CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluPro
 3422 TGCCCCCCCCGACTCCGACGCTGAGTCCTATTCCCTCATGCCCGGGCTGGAGGGGGAGCCT
 ACGGGGGGGCTGAGGCTGCCACTCAGGATAAGGAGGTACGGGGGGACCTCCCCCTCGGA
 ^
 3437 EAM11051,
 · GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTATGGTCAACGGTCAGTAGTGAGGCCAACGCGGGAG
 CCCCTAGGCCTAGAATCGCTGCCAGTACCGTACAGTCCGTTGCGCCT
 ^ ^
 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
 · AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla
 3542 GATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCGTGCGCC
 CTACAGCACACGACGAGTTACAGAATGAGAACCTGCGCTGAGCAGTGGGCACGCGG
 ^
 3589 DRA3, 3600 SAC2,
 · AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
 3602 GCGGAAGAACAGAAAATGCCCATCAATGCACTAAGCAACTCGTTGCTACGTACCCACAAT
 CGCCTTCTTGCTTTGACGGTAGTTACGTGATTGAGCAACGATGCGAGTGGTGT
 ^
 3611 ALWN1, 3655 PFLM1,
 · LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
 3662 TTGGTGTATTCCACCACCTCACGCACTGCTTGCCTACGGCAGAACGAAAGTCACATTGAC
 AACACACATAAGGTGGTAGTGCACGAAAGTTCCGTCTTCAGTGTAAACTG
 ^
 3681 DRA3,
 · ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
 3722 AGACTGCAAGTCTGGACAGCCATTACCAAGGACGTACTCAAGGAGGTAAAGCAGCGCG
 TCTGACGTTCAAGACCTGTCGTAATGGTCTGCATGAGTTCTCCAATTGCGCCGC

FIGURE 14 - Page 7

SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
 3762 TCAAAAGTGAAGGCTAACCTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCACAC
 AGTTTCACTCCGATTGACCGATAGGCATCTCCTCGAACGTCGGACTGCGGGGTGTG
 3816 Hind3,
 SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
 3842 TCAGCCAAATCCAAGTTGGTTATGGGGCAAAAGACGTCGCTGCATGCCAGAAAGGCC
 AGTCGGTTAGGTTCAAACCAATACCCGTTCTGCAGGCAACGGTACGGTCTTCCGG
 3875 AAT2, 3890 BGLI,
 ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp
 3902 GTAAACCCACATCAAACCTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATAGAC
 CATTGGGTGTAGTTGAGGCACACCTTCTGGAAGACCTTCTGTTACATTGTGGTTATCTG
 ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyArgLys
 3962 ACTACCACATCATGGCTAAGAACGAGGTTCTGCGTCAGCCTGAGAAGGGGGTCGTAAG
 TGATGGTAGTACCGATTCTGCTCCAAAGACGCAAGTCGGACTCTCCCCCAGCATTC
 ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
 4022 CCAGCTCGTCTCATCGTGTCCCCGATCTGGCGTGCCTGCGAAAGATGGCTTG
 GGTCGAGCAGAGTAGCACAAAGGGCTAGACCCGCACGCGCACACGCTTTCTACCGAAC
 TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
 4082 TACGACGTGGTTACAAAGCTCCCCTGGCGTGATGGGAAGCTCCTACGGATTCCAATAC
 ATGCTGCACCAATGTTGAGGGAACCGGCACTACCCCTCGAGGATGCCTAAGGTTATG
 SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
 4142 TCACCAGGACAGCGGGTTGAATTCTCGTGCAGCGTGGAAAGTCCAAGAAAACCCCAATG
 AGTGGCTCTGCGCCCAACTTAAGGAGCACGTTGCACCTTCAGGTTCTTGGGTTAC
 4160 EORI,
 GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
 4202 GGGTTCTCGTATGATAACCGCTGCTTGACTCCACAGTCACTGAGAGCGACATCCGTACG
 CCCAAGAGCATACTATGGCGACGAAACTGAGGTGTCAGTACTCTCGTGTAGGCATGC
 4229 DRD1, 4236 ALWN1,
 GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
 4262 GAGGAGGCAACTACCAATGTTGTGACCTCGACCCCCCAAGCCCCGTGGCCATCAAGTCC
 CTCCCTCGTTAGATGGTTACAACACTGGAGCTGGGGTTCGGGCGACCGGTAGTTCAAG
 4301 BGLI, 4308 BALI,
 LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
 4322 CTCACCGAGAGGCTTATGTTGGGGCCCTTACCAATTCAAGGGGGAGAACTGCGGC
 GAGTGGCTCTCGAAATACAACCCCGGGAGAATGGTTAAGTCCCCCTCTTGACGCCG
 4345 APAI,
 TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
 4382 TATCGCAGGTGCCGCGCAGCGCGTACTGACAACAGCTGTGGTAACACCCTCACTTGC
 ATAGCGTCCACGGCGCGCTCGCCGATGACTGTTGATCGACACCATTGTGGAGTGAACG

FIGURE 14 - Page 8

TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
 4442 TACATCAAGGCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTG
 ATGTAGTCCGGCCCCGTGGACAGCTCGGCCTCCAGGTCCTGACGTGGTACGAGCAC
 ^
 4452 SMA1 XMA1,
 CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
 4502 TGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGTCCAGGAGGACGCCGGCAGC
 ACACCGCTGCTGAATCAGCAATAGACACTTCGCGCCCCCAGSTCCTGCGCCGCTCG
 ^
 4508 DRD1, 4511 TTH3I,
 LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGin
 4562 CTGAGAGCCTTCACGGAGGTATGACCAGGTACTCCGCCCCCCTGGGGACCCCCCACAA
 GACTCTCGGAAGTGCCTCCGATACTGGTCATGAGGCGGGGGGACCCCTGGGGGTGTT
 ^
 ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 4622 CCAGAAATACGACTTGGAGCTCATAAACATCATGCTCCTCCAACGTGTCAGTCGCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTG
 ^
 4637 SAC1,
 GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
 4682 GGCCTGGAAAGAGGGTCACTACCTCACCCGTGACCCCTACAACCCCCCTCGCGAGAGCT
 CGCGACCTTCTCCAGATGATGGAGTGGGCACTGGATGTTGGGGGAGCGCTCTCGA
 ^
 4731 NRUI,
 AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 4742 GCGTGGGAGACAGCAAGACACACTCCAGTCATTCTGGCTAGGCAACATAATCATGTT
 CGCACCCCTCTGCGTTCTGAGGTCAGTTAAGGACCGATCCGTTGATTAGTACAAA
 ^
 AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 4802 GCCCCCACACTGTGGCGAGGATGATACTGATGACCCATTCTTAGCGTCCTTATAGCC
 CGGGGGTGTGACACCCGCTCTACTATGACTACTGGTAAAGAAATCGCAGGAATATCGG
 ^
 4806 PFLM1, 4807 DRA3,
 ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu
 4862 AGGGACCAAGCTTGAACAGGCCCTGATTGCGAGATCTACGGGCCTGCTACTCCATAGAA
 TCCCTGGTCGAATTGTCCGGAGCTAACGCTCTAGATGCCCGGACGATGAGGTATCTT
 ^
 4893 BGL2,
 ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
 4922 CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTCACTCCAC
 GGTGACCTAGATGGAGGTAGTAAGTTCTGAGGTACCGGAGTCGCGTAAAGTGAGGTG
 ^
 4954 NCOI,
 SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
 4982 AGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAAACTGGGGTACCG
 TCAATGAGAGGTCCACTTTAGTTATCCCACCGGCGTACGGAGTCTTGTAAACCCATGGC
 ^
 5015 SPHI, 5035 KPNI,
 ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly

FIGURE 14 - Page 9

5042 CCCTTGCAGCTTGGAGACACCGGGCCGGAGCGTCCGCGTAGGCTCTGCCAGAGGA
GGAAACGCTCGAACCTCTGTGGCCCGGCCTCGCAGGCAGATCCGAAGACCAGGCTCCT
5064 APAI, 5091 BALI,
GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
5102 GGCAGGGCTGCCATATGTGCCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTAAA
CCGTCGGACGGTATAACACCGTTATGGAGAAGTTGACCCGTCAATTCTTGTTCGAGTT
5113 NDEI,
LeuThrProIleAlaAlaAlaGlyGlnLeuAspLeuSerGlyTrpPheThrAlaGlyTyr
5162 CTCACTCCAATAGCGGGCGCTGGCCAGCTGGACTTGTCCGGCTGGTTCACGGCTGGCTAC
GAGTGAGGTTATCGCCGGGACCGGTGACCTGAACAGGCCGACCAAGTGGCGACCGATG
5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,
SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
5222 AGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCGCTGGATCTGGTTTGC
TCGCCCCCTGTAAATAGTGTGCGCACAGAGTACGGGCCGGCGACCTAGACCAAAACG
5240 DRA3,
LeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgOP
5282 CTACTCCTGCTTGCTGCAGGGGTAGGCATCTACCTCCTCCCAACCGATGAATAGTCGAC
GATGAGGACGAACGACGTCCCCATCCGTAGATGGAGGAGGGTTGGCTACTTATCAGCTG
5295 PSTI, 5336 SALI,

FIGURE 15



FIGURE 16 - Page 1

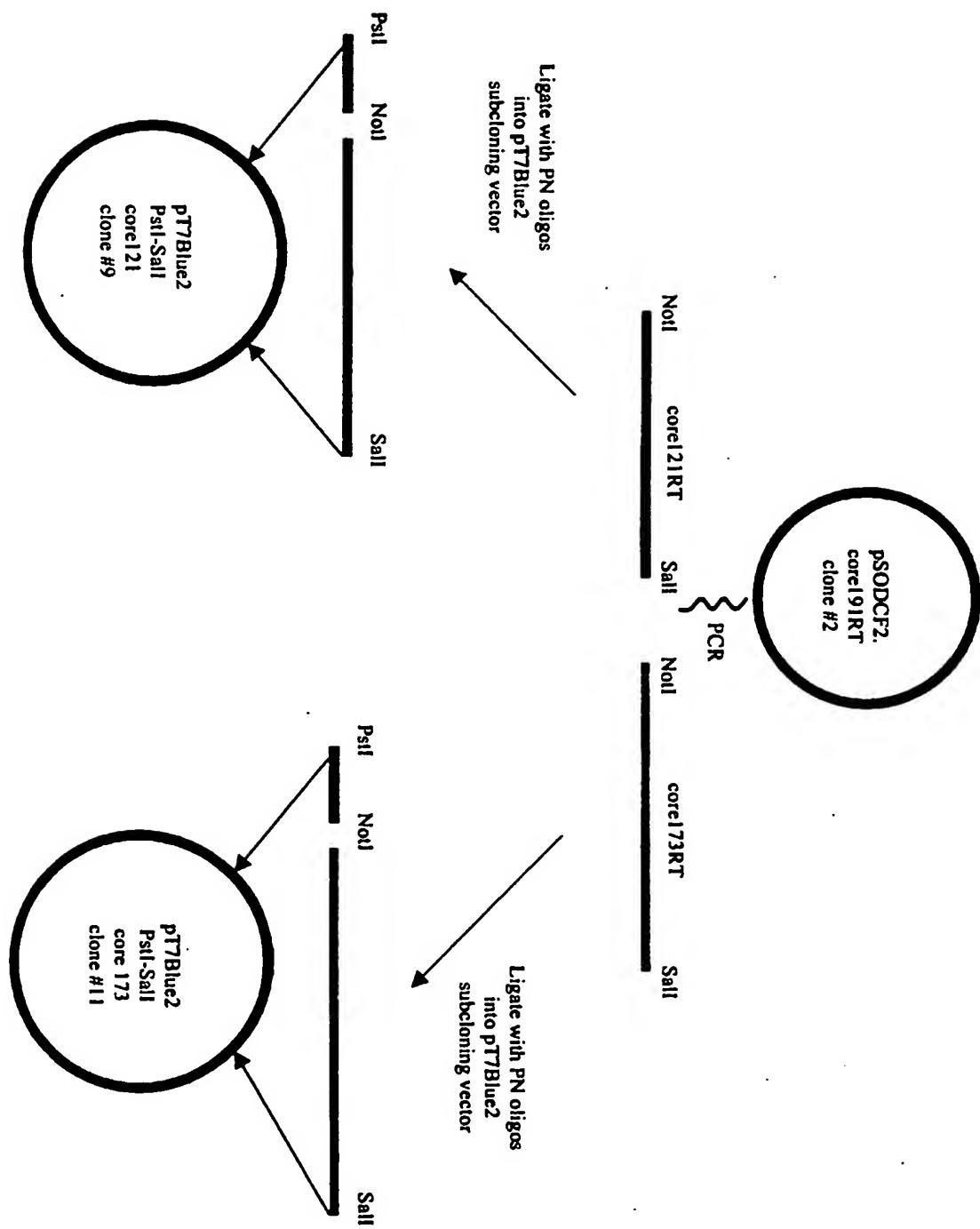


FIGURE 16 - Part 2

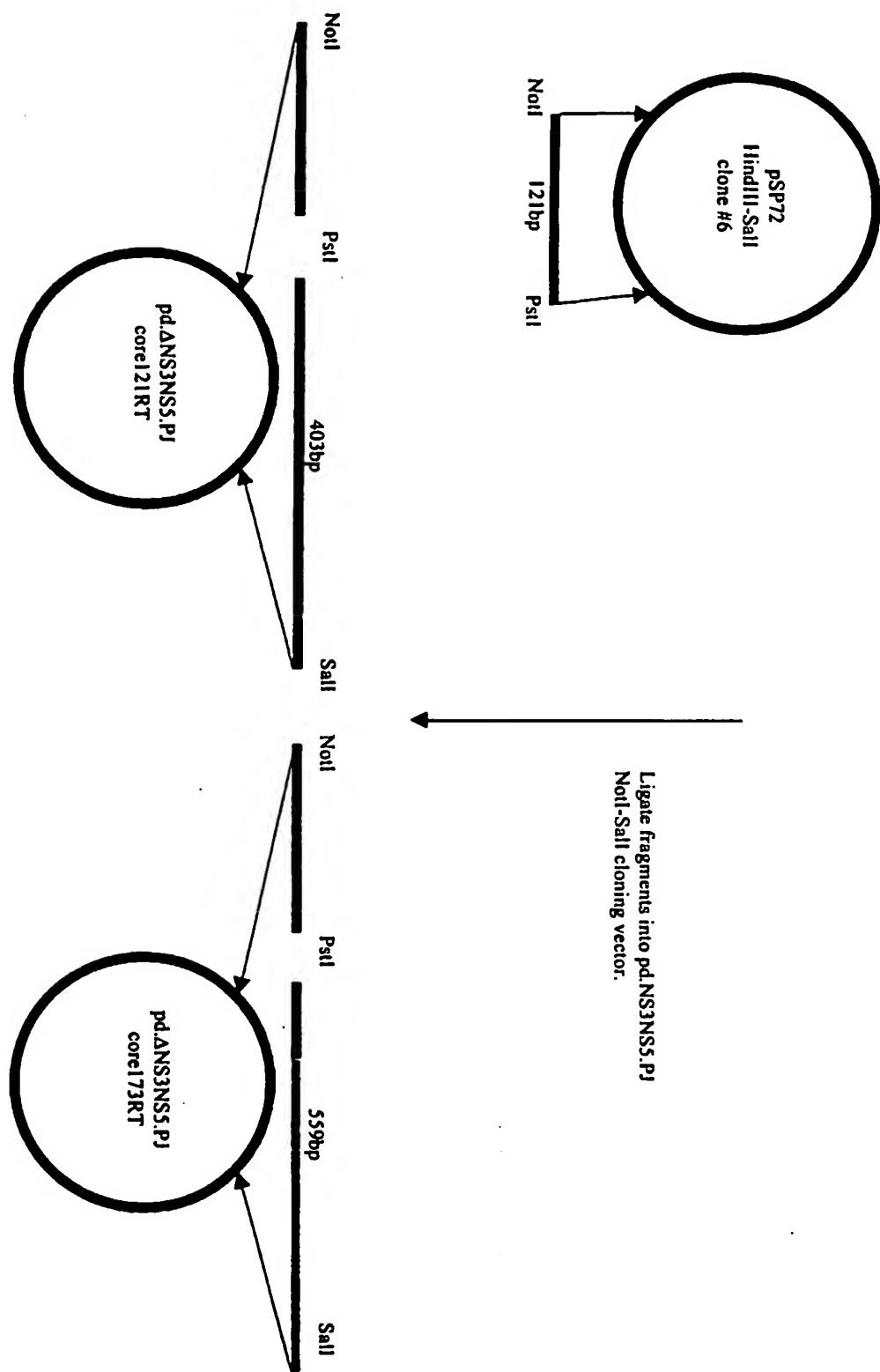


FIGURE 17 - Page 1

MetAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsn

2 AGCTTACAAAACAAATGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAAC
 TCGAATGTTTGTTCACGACGTACGTCGAGTCCGATATTCCACGATCATGAGTTG
 ^ ^ ^
 1 HIND3, 24 NDEI, 52 SCAI,

ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAsp
 62 CCCTCTGTTGCTGCAACACTGGGCTTGGTGCCTACATGTCCAAGGCTCATGGGATCGAT
 GGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCCTAGCTA
 ^
 116 CLAI,

ProAsnIleArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSerThr
 122 CCTAACATCAGGACCGGGGTGAGAACAAATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTTGTAAATGGTGACCGTCGGGTAGTGCATGAGGTGG

TyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCys
 182 TACGGCAAGTTCTTGCCGACGGCGGGTGCCTCGGGGGCGCTTATGACATAATAATTGT
 ATGCCGTTCAAGGAACGGCTGCCGCCACGAGCCCCCGCGAATACTGTATTATTAAACA

AspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGln
 242 GACGAGTGCCTCCACGGATGCCACATCCATCTGGCATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTGACAGGAACGGTT

AlaGluThrAlaGlyAlaArgLeuValLeuAlaThrAlaThrProProGlySerVal
 302 GCAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCGCTCTGACCAACACGAGCGGTGGCGTGGGGAGGCCGAGGCAG
 ^
 303 ALWN1,

ThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPhe
 362 ACTGTGCCCATCCAAACATCGAGGAGGTGCTCTGTCACCACCGGAGAGATCCCTTT
 TGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAA

TyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHis
 422 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGAGACATCTCATCTCTGTCA
 ATGCCGTTCCGATAGGGGGAGCTTCATTAGTCCCCCTCTGTAGAGTAGAACAGTA

FIGURE 17 - Page 2

SerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaVal
 482 TCAAAAGAAGAAGTGCACGAACTCGCCGAAAGCTGGTCGATTGGCATCAATGCCGTG
 AGTTTCTCTTCACGCTGCTTGAGCGGCCTTCGACCAGCGTAACCGTAGTTACGGCAC

 AlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValVal
 542 GCCTACTACCGCGGTCTTGACGTGTCGTCATCCCGACCAGCGGCATGTTGTCGTCGTG
 CGGATGATGGCGCCAGAACAGCACAGCAGTAGGGCTGGTCGCCCTACAACAGCAGCAC

 550 SAC2, 560 DRD1,

 AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsn
 602 GCAACCCGATGCCCTCATGACC GGCTATACCGGCAGCTCGACTCGGTGATAGACTGCAAT
 CGTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACGTTA

 615 BSPH1,

 ThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIle
 662 ACGTGTGTCACCCAGACAGTCGATTCAGCCTTGACCTTACCTCACATTGAGACAATC
 TGCACACAGTGGGCTGTCAGCTAAAGTCGGAACCTGGGATGGAAGTGGTAACCTGTTAG

 ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 722 ACGCTCCCCAAGATGCTGTCTCCCGACTCAACGTCGGGGCAGGACTGGCAGGGGGAAAG
 TGCAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCCTTC

 ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 782 CCAGGCATCTACAGATTGTGGCACCGGGGGAGCGCCCTCCGGCATGTTGACTCGTCC
 GGTCCCGTAGATGTCTAACACCGTGGCCCCCTCGCGGGGAGGCCGTACAAGCTGAGCAGG

 816 BGLI, 833 DRD1,

 ValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThr
 842 GTCCTCTGTGAGTGTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCGCCGAGACT
 CAGGAGACACTCACGATACTGGTCCGACACGAAACCATACTCGAGTGCAGGGGGCTCTGA

 881 SAC1,

 ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 902 ACAGTTAGGCTACGAGCGTACATGAACACCCGGGGCTTCCCGTGTGCCAGGACCATCTT
 TGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGCTGGTAGAA

 931 SMA1 XMA1,

 GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln
 962 GAATTTGGGAGGGCGTCTTACAGGCCTCACTCATATAGATGCCACTTCTATCCAG
 CTTAAAACCTCCCGAGAAATGTCCGGAGTGAGTATCTACGGGTGAAAGATAGGGTC

 985 STUI,

 ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
 1022 ACAAAAGCAGAGTGGGGAGAACCTCCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
 TGTTTCGTCTCACCCCTCTTGGAAAGGAATGGACCATCGCATGGTCGGTGGCACACGCGA

 1069 DRA3,

 ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLys
 1082 AGGGCTCAAGCCCCCTCCCCCATCGTGGGACCAGATGTGAAAGTGGATTCCGCTCAAG

FIGURE 17 - Page 3

TCCCGAGTTGGGGAGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTC

1142 ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 CCCACCCCTCCATGGGCCAACACCCCTGCTATACAGACTGGCGCTGTTCAAGATGAAATC
 GGGTGGGAGGTACCCGGTTGTGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAG

1150 NCOI,

1202 ThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluVal
 ACCCTGACGCACCCAGTCACCAAATACATCATGACATGCTCGGCCGACCTGGAGGTC
 TGGGACTGCGTGGGTCAAGTGGTTATGTAGTAGTACTGTACGTACAGCCGGCTGGACCTCCAG

1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,

1262 ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 GTCACGAGCACCTGGGTGCTCGTGGCGCCTGGCTGCTTGGCCGCGTATTGCCTG
 CAGTGCTCGTGGACCCAGCAGCAACGCCGCAGGACCGACGAAACCGGCGCATAACGGAC

1322 SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIle
 TCAACAGGCTGCGTGGTCATAGTGGCAGGGTCGTCTGTCCGGGAAGCCGGCAATCATA
 AGTTGTCCGACGCACCACTACCCGTCAGCAGAACAGGCCCTCGGCCGTTAGTAT

1369 NAEI,

1382 ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
 CCTGACAGGGAAAGTCCTCTACCGAGAGTCGATGAGATGAAAGAGTGCTCTCAGCACTTA
 GGACTGTCCCCTCAGGAGATGGCTCTCAAGCTACTCTACCTCTCACGAGAGTCGTGAAT

1385 DRD1,

1442 ProTyrIleGluGlnGlyMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 CCGTACATCGAGCCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAAGGCCCTCGGCCCTC
 GGCATGTAGCTCGTCCCTACTACGAGCGGCTCGTCAAGTTCTCCGGGAGCCGGAG

1502 LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 CTGCAGACCGCGTCCCGTCAGGCAGAGTTATGCCCTGCTGTCAGACCAACTGGCAA
 GACGTCTGGCGCAGGGCAGTCGTCTCAAAGCAGGTCTGGTTGACCGTT

1502 PSTI, 1507 TTH3I,

1562 LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
 AAAACTCGAGACCTCTGGGCGAAGCATATGTGGAACCTTCATCAGTGGATAACAATACTTG
 TTTGAGCTCTGGAAGACCCGCTCGTACACCTTGAAGTAGTCACCCCTATGTTATGAAC

1565 XHOI, 1586 NDEI,

1622 AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla
 CGGGGCTTGTCAACGCTGCCGTAAACCCGCCATTGCTTCATTGATGGCTTTACAGCT
 CGCCCCAACAGTTCGACGGACCATGGCGGTAACGAAGTAACCGAAAATGTCGA

1643 BSTE2, 1677 ALWN1 PVU2,

1682 AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp
 GCTGTCACCAGCCCCACTAACCAACTAGCAGAAACCCCTCTTCACATATTGGGGGGGTGG
 CGACAGTGGTCGGGTGATTGGTATCGGTTGGGAGGAGAAGTTGTATAACCCCCCCCACC

FIGURE 17 - Page 4

1742 ValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAla
 GTGGCTGCCAGCTGCCGCCCTCGGTGCCGCTACTGCCTTGTGGCGCTGGCTTAGCT
 CACCGACGGGTCGAGCGGGGGCCACGGCGATGACGGAAACACCCGCGACCGAATCGA

1794 ESP1,

1802 GlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyr
 GGCGCCGCCATCGGCAGTGTGGACTGGGAAGGTCTCATAGACATCCTGCAGGGTAT
 CGCGGGCGGTAGCGTCACAACCTGACCCCCTCAGGAGTATCTGTAGGAACGTCCCATA
 1802 KAS1 NARI,

1862 GlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSer
 GGC CGGGCGTGGCGGGAGCTCTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCCTCC
 CGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTAGTACTCGCCACTCCAGGGGAGG

1878 SAC1, 1899 BSPH1,

1922 ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly
 ACGGAGGACCTGGTCAATCTACTGCCGCATCCTCTGCCCGAGCCCTCGTAGTCGGC
 TGCCTCCTGGACCAGTTAGATGACGGCGGTAGGAGAGCGGGCCTGGAGCATCAGCCG
 1928 TTH3I,

1982 ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
 GTGGCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCAGGGGGCAGTGCAGTGG
 CACCAAGACACGTCGTTATGACGCGGCCGTGCAACCAGGGCCGCTCCCCCGTCACGTCACC
 2004 NAEI, 2017 SMAI XMAI,

2042 MetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
 ATGAACCGGCTGATAGCCTCGCCTCCGGGGAAACCATGTTCCCCCACGCACTACGTG
 TACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGGGTGCCTGATGCAC
 2067 SMAI XMAI, 2093 DRA3,

2102 ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
 CCGGAGAGCGATGCAGCTGCCCGCCTACTGCCATACTCAGCAGCCTCAGTAAACCCAG
 GGCCCTCGCTACGTCGACGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGTC
 2115 PVU2, 2159 ALWN1,

2162 LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
 CTCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCC
 GAGGACTCCGCTGACGTGGTCACCTATTGAGCCTCACATGGTGGAGGTACGAGGCCAAGG
 2164 MST2, 2220 ECON1,

2222 TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
 TGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTA
 ACCGATTCCCTGTAGACCTGACCTATACGCTCCACAATCGCTGAAATTCTGGACCGAT

2282 LysAlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyr
 AAAGCTAACGCTCATGCCACAGCTGCCCTGGATCCCCTTGTGTCCTGCCAGCGCGGGTAT
 TTTCGATTGAGTACGGTGTGACGGACCCCTAGGGAAACACAGGACGGTCGCGCCCCATA
 2285 ESP1, 2300 PVU2, 2310 BAMHI,

FIGURE 17 - Page 5

2342 LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 AAGGGGGTCTGGCGAGGGGA^CGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATC
 TTCCCCAGACCGCTCCCCTGCCTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAG

2402 ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 ACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCCTGTAC

2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,

2462 TrpSerGlyThrPheProIleAsnAlaTyrThrGlyProCysThrProLeuProAla
 TGGAGTGGGACCTTCCCCATTAAATGCCTACACCACGGGCCCCCTGTACCCCCCTTCCTGCG
 ACCTCACCTGGAAAGGGTAATTACGGATGTGGTCCCCGGGACATGGGGGAAGGACGC

2480 ASE1, 2497 APA1,

2522 ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln
 CCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCAG
 GGCTTGATGTGCAAGCGCGATACTCCACAGACGTCTCCTATGCACCTCTATTCCGTC

2553 PSTI,

2582 ValGlyAspPheHisTyrValThrGlyMetThrAspAsnLeuLysCysProCysGln
 GTGGGGGACTCCACTACGTGACGGGTATGACTACTGACAATCTTAAATGCCGTGCCAG
 CACCCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTACGGGCACGGTC

2594 DRA3,

2642 ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro
 GTCCCATGCCCGAATTTTACAGAAATTGGACGGGTGCCCTACATAGTTGCC
 CAGGGTAGCGGGCTAAAAAGTGTCTAACCTGCCACGCGGATGTATCAAACGCGGG

2702 ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTTGCTGCCGGAGGAGGTATTCAGAGTAGGACTCCACGAATACCG
 GGGACGTTGGAACGACGCCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGC

2757 HGIE2,

2762 ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
 GTAGGGTCGCAATTACCTTGCAGGCCAACCGGACGTGGCGTGTGACGTCCATGCTC
 CATCCCAGCGTTATGGAACGCTCGGGCTTGGCCTGCACCGCACAACTGCAGGTACGAG

2809 AAT2,

2822 ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
 ACTGATCCCTCCCATATAACAGCAGAGGCAGGCCGGCGAAGGTTGGCGAGGGGATCACCC
 TGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCGTTCCAACCGCTCCCTAGTGGG

2850 EAG1 XMA3,

2882 ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
 CCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCAAGGCAACTTGC
 GGGAGACACCGGTCGAGGAGGCCATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACG

2889 BALI, 2903 NHEI.

FIGURE 17 - Page 6

2942 ThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGln
 ACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAG
 TGGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGGTGGAGGATACCTCCGTC
 ^ ^
 2966 ESP1, 2969 SACI,
 GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 3002 GAGATGGGCGGCAACATCACCAAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCC
 CTCTACCCGCCGTGAGTGGTCCCAACTCAGTCTTTGTTTACCAACTAAGACCTGAGG
 PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 3062 TTGATCCCGCTTGTGGCGGAGGAGGACGAGCAGGAGATCTCGTACCCGAGAAATCCTG
 AAGCTAGGCGAACACCCTCCCTGCTGCCCTAGAGGCATGGCGTCTTAGGAC
 ^
 3096 BGL2,
 ArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro
 3122 CGGAAGTCTGGAGATTGCCCTAGGCCCTGCCGTTGGCGCGCCGACTATAACCC
 GCCTTCAGAGCCTCTAACGGGTCCGGACGGCAAACCCGCGCCGCGCTGATATTGGGG
 ^ ^
 3143 ALWN1, 3164 EAG1 XMA3,
 ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCCTAGTGGAGACGTTGGAAAAAGCCCGACTACGAACCACCTGTTGGTCCATGGCTGCCCG
 GGCATCACCTCTGCACCTTTCTGGCTGATGCTTGGTGGACACCAGGTACCGACGGGC
 ^ ^
 3217 HGIE2, 3229 NCOI,
 LeuProProProLysSerProProValProProProArgLysLysArgThrValValLeu
 3242 CTCCACCTCAAAGTCCCCTCTGTGCCTCCGCCTGGAAAGAAGCGGACGGTGGTC
 GAAGGTGGAGGTTTCAGGGAGGACACGGAGGGAGCCTTCGCCTGCCACCAGGAG
 ^ ^
 ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer
 3302 ACTGAATCAACCCTATCTACTGCCTGGCCGAGCTGCCACCAGAAGCTTGGCAGCTCC
 TGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGG
 ^ ^
 3332 SACI, 3346 HIND3,
 SerThrSerGlyIleThrGlyAspAsnThrThrSerSerGluProAlaProSerGly
 3362 TCAACTTCCGGATTACGGGCACAAATACGACAACATCCTCTGAGCCGCCCTCTGGC
 AGTTGAAGGCCGTAATGCCGCTGTATGCTGTTAGGAGACTCGGGCGGGAAAGACCG
 ^
 CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluPro
 3422 TGCCCCCCCCGACTCGACGCTGAGTCCTATTCCCTCATGCCCGGGCTGGAGGGGGAGCCT
 ACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGACCTCCCCCTCGGA
 ^
 3437 EAM11051,
 GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAG
 CCCCTAGGCCTAGAATCGCTGCCAGTACCAAGTGGCCAGTCATCACTCCGGTTGCGCCTC
 ^ ^ ^
 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
 AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla
 3542 GATGTCGTGCTGCTCAATGTCTTACTCTTGACAGGCGCACTCGTCACCCGTGCGCC
 CTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGGAGCAGTGGGGCACCGCG

FIGURE 17 - Page 7

3589 DRA3, 3600 SAC2,

3602 AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
GCGGAAGAACAGAAACTGCCATCAATGCACTAAGCAACTCGTGTACGTCACCAAAAT
CGCCTTCTTGCTTTGACGGTAGTTACGTGATTGAGCAACGATGCAGTGGTGTAA

3611 ALWN1, 3655 PFLM1,

3662 LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
TTGGTGTATTCCACCACCTCACGCAGTGTGCTGCCAAAGGCAGAAGAAAGTCACATTGAC
AACCACATAAGGTGGTGGAGTGCACGAAACGGTTCCGTCTTCAGTGTAAACTG

3681 DRA3,

3722 ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
AGACTGCAAGTCTGGACAGCCATTACCAAGGACGTACTCAAGGAGGTTAAAGCAGCGCG
TCTGACGTTCAAGACCTGCGTAATGGTCCTGCAATGAGTTCTCCAATTTCGTCGCCGC

3782 SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
TCAAAAGTGAAGGCTAACTGCTATCCGTAGAGGAAGCTGAGCCTGACGCCACAC
AGTTTCACTTCCGATTGAACGATAGGCATCTCCTCGAACGTCGGACTGCGGGGTGTG

3816 HIND3,

3842 SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
TCAGCCAAATCCAAGTTGGTTATGGGGAAAAAGACGTCCGTGCCATGCCAGAAAGGCC
AGTCGGTTAGGTTCAAACCAATACCCGTTTCTGCAGGCAACGGTACGGTCTTCGG

3875 AAT2, 3890 BGLI,

3902 ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp
GTAACCCACATCAACTCCGTGTGGAAAGACCTCTGGAAAGACAATGTAACACCAATAGAC
CATGGGTGTAGTTGAGGCACACCTTCTGGAAAGACCTCTGTTACATTGTGGTTATCTG

3962 ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArgLys
ACTACCACATGGCTAAAGAACGAGGTTTCTGCCTCAGCCTGAGAAGGGGGTCGTAAG
TGATGGTAGTACCGATTCTGCTCCAAAGACGCAAGTCGGACTCTCCCCCAGCATTC

4022 ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
CCAGCTCGTCTCATCGTGTCCCGATCTGGCGTGCACGTGCGAAAGATGGCTTG
GGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGACACGCGCACACGTTTCTACCGAAC

4082 TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
TACGACGTGGTTACAAAGCTCCCTTGGCGTGATGGGAAGCTCCTACGGATTCCAATAC
ATGCTGCACCAATGTTGAGGGGAAACGGCACTACCCCTCGAGGATGCCTAAGGTTATG

4142 SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
TCACCAGGACAGCGGGTTGAATTCTCGTGCAGCGTGGAAAGTCCAAGAAAACCCCAATG
AGTGGTCTGTCGCCAACCTAAGGAGCACGTTGCACCTCAGGTTCTTGGGTAC

4160 ECORI,

4202 GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
GGTTCTCGTATGATACCCGCTGCTTGACTCCACAGTCAGTGAGAGCGACATCCGTACG
CCCAAGAGCATACTATGGCGACGAAACTGAGGTGTCAGTGAATCTCGCTGTAGGCATGC

FIGURE 17 - Page 8

4229 DRD1, 4236 ALWN1,

4262 GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
 GAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGTGCCCATCAAGTCC
 CTCCTCCGGTAGATGGTTACAACACTGGAGCTGGGGTTCGGCGCACCGTAGTCAGG

4301 BGLI, 4308 BALI,

4322 LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
 CTCACCGAGAGGCTTTATGTTGGGGCCCTCTTACCAATTCAAGGGGGAGAACTGCGGC
 GAGTGGCTCTCCGAAATAACACCCCCGGGAGAATGGTTAACGTTCCCCCTTGACGCCG

4345 APA1,

4382 TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
 TATCGCAGGTGCCGCGCAGCGCGTACTGACAACTAGCTGTGGTAACACCCACTTGC
 ATAGCGTCCACGGCGCGCTCGCCGATGACTGTTGATCGACACCATTGTGGAGTGAACG

4442 TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
 TACATCAAGGCCGGCAGCCTGTCGAGCCGAGGGCTCCAGGACTGCACCATGCTCGTG
 ATGTAGTTCCGGGCCGTGGACAGCTGGCGTCCGAGGTCTGACGTGGTACGAGCAC

4452 SMA1 XMA1,

4502 CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
 TGTGGCGACGACTTAGTCGTTATCTGAAAGCGCGGGGTCCAGGAGGACGCGGCGAGC
 ACACCGCTGCTGAATCAGCAATAGACACTTCGCGCCCCCAGGTCTCTGCGCCGCTCG

4508 DRD1, 4511 TTH3I,

4562 LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
 CTGAGAGCCTCACGGAGGCTATGACCAGGTACTCCGCCCCCTGGGACCCCCCACAA
 GACTCTCGGAAGTGCCTCGATACTGGCCATGAGGCGGGGGGGACCCCTGGGGGTGTT

4622 ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 CCAGAACGACTTGGAGCTATAACATCATGCTCTCCAACGTGTCAGTCGCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTGCACAGTCAGCGGGTGCTG

4637 SAC1,

4682 GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
 GGCGCTGGAAAGAGGGTCAACTACCTCACCGTGACCCCTACAAACCCCTCGCAGAGCT
 CCGCGACCTTCTCCAGATGATGGAGTGGCACTGGATGTTGGGGAGCGCTCTCGA

4731 NRUI,

4742 AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 GCGTGGGAGACAGCAAGACACACTCCAGTCAATTCTGGCTAGGCAACATAATCATGTTT
 CGCACCCCTCTGCGTTCTGTGAGGTCAAGTAAGGACCGATCCGTTGTATTAGTACAAA

4802 AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 GCCCCCACACTGTGGCGAGGATGATACTGATGACCCATTCTTAGCGTCCTTATAGCC
 CGGGGGTGTGACACCCGCTCTACTATGACTACTGGTAAAGAAATCGCAGGAATATCGG

4806 PFLM1, 4807 DRA3,

ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu

FIGURE 17 - Page 9

4862 AGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGCCTGCTACTCCATAGAA
 TCCCTGGTCGAACCTGTCCGGAGCTAACGCTCTAGATGCCCGGACGATGAGGTATCTT
 ^
 4893 BGL2,
 . ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
 4922 CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTCACTCCAC
 GGTGACCTAGATGGAGGTTAGTAAGTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTG
 ^
 4954 NCOI,
 . SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
 4982 AGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAATGGGTACCG
 TCAATGAGAGGTCCACTTAGTTATCCCACCGGCGTACGGAGTCTTTGAACCCCATGGC
 ^
 5015 SPHI, 5035 KPNI,
 . ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly
 5042 CCCTTGCAGCTTGGAGACACCGGGCCCGAGCGTCCCGCTAGGCTTCTGCCAGAGGA
 GGGAACGCTCGAACCTCTGTGGCCCGGCCCTCGCAGGCGCATCCGAAGACCGGTCTCCT
 ^
 5064 APAI, 5091 BALI,
 . GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
 5102 GGCAGGGCTGCCATATGTGGCAAGTACCTCTCAACTGGCAGTAAGAACAAAGCTCAA
 CCGTCCCGACGGTATAACACCGTTATGGAGAAGTTGACCCGTATTCTGTTGAGTT
 ^
 5113 NDEI,
 . LeuThrProIleAlaAlaAlaGlyGlnLeuAspLeuSerGlyTrpPheThrAlaGlyTyr
 5162 CTCACTCCAATAGCGGCCGCTGGCAGCTGGACTTGTCCGGCTGGTTCACGGCTGGCTAC
 GAGTGAGGTTATCGCCGGCACCAGGTGACCTGAACAGGCCGACCAAGTGGCGACCGATG
 ^ ^ ^
 5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,
 . SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
 5222 AGCGGGGGAGACATTATCACAGCGTGTCTCATGCCCGCCCCGCTGGATCTGGTTTGC
 TCGCCCCCTCTGTAAATAGTGTGCGACAGAGTACGGGCCGGGCGACCTAGACCAAAACG
 ^
 5240 DRA3,
 . LeuLeuLeuAlaAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgMetSerThrAsn
 5282 CTACTCCTGTTGCTGCAGGGTAGGCATCTACCTCTCCCAACCGAATGAGCACGAAT
 GATGAGGACGAACGACGTCCCCATCCGTAGATGGAGGAGGGTTGGCTACTCGTGCTTA
 ^
 5295 PSTI,
 . ProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGlnAspValLysPhe
 5342 CCTAAACCTCAAAGAAAGACCAAACGTAACACCAACCGGCCGCAGGACGTCAAGTTC
 GGATTGGAGTTCTTCTGGTTGCATTGTGGTGGCCGCCGGCGTCTGCAGTTCAAG
 ^ ^ ^
 5380 NOTI, 5381 EAG1 XMA3, 5390 AAT2, 5401 SMA1 XMAI,
 . ProGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeu
 5402 CGGGGTGGCGGTCAAGATGTTGGAGTTACTTGTGCGCGCAGGGGCCCTAGATTG
 GGCCCCACCGCCAGTCTAGCAACCACCTCAAATGAACAAACGGCGCTCCCCGGGATCTAAC
 ^

FIGURE 17 - Page 10

5449 APAI,

5462 GlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnPro
GGTGTGCGCGCAGAGAAAGACTTCCGAGCGGTGCAACCTCGAGGTAGACGTAGCCT
CCACACGCGCGCTGCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCAGTCGGA

5467 BSSH2, 5478 XMNI, 5502 XHOI, 5511 AAT2,

5522 IleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProT:pPro
ATCCCCAAGGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCAGGTACCCCTGGCC
TAGGGGTTCCGAGCAGCCGGCTCCCGTCCGGACCCGAGTCGGGCCATGGAACCGGG

5548 ALWN1, 5558 ESP1, 5564 SMAI XMAI, 5568 KPNI,

5582 LeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArg
CTCTATGGCAATGAGGGCTGCGGGTGGCGGGATGGCTCTGTCTCCCGTGGCTTCGG
GAGATACCGTTACTCCGACGCCAACCGCCCTACCGAGGACAGAGGGCACCGAGAGCC

5642 ProSerTrpGlyProThrAspProArgArgSerArgAsnLeuGlyLysOC AM
CCTAGCTGGGGCCCCACAGACCCCCCGGCGTAGGTGCGCGCAATTGGGTAAGTAATAGTCG
GGATCGACCCCCGGGTGTCTGGGGCCGCATCCAGCGCGTTAAACCCATTCAATTACGC

5650 APAI, 5698 SALI,

5702 AC
TG

FIGURE 18 - Page 1

MetAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsn
 2 AGCTTACAAAACAAAATGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTAGTACTCAAC
 TCGAATGTTTGTGTTACCGACGTATACGTCGAGTCCCGATATTCCACGATCATGAGTTG
 ^ .
 1 HIND3, 24 NDEI, 52 SCAI,

 ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAsp
 62 CCCTCTGGCTGCAACACTGGGCTTGGCTTACATGTCCAAGGCTATGGGATCGAT
 GGGAGACAAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCAGCTA

 116 CLAI,

 ProAsnIleArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSerThr
 122 CCTAACATCAGGACCGGGGTGAGAACATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCACCTCTGTTAATGGTGACCGTCGGGTAGTGCATGAGGTGG

 TyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCys
 182 TACGGCAAGTCCCTTGCCGACGGCGGGTGCTCGGGGGCGCTATGACATAATAATTG
 ATGCCGTTCAAGGAACGGCTGCCACGAGCCCCCGCGAATACTGTATTATTAAACA

 AspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGln
 242 GACGAGTGCACACTCCACGGATGCCACATCCATCTGGCATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTAGGGTAAACCGTAACCGTGACAGGAACGGTT

 AlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGlySerVal
 302 GCAGAGACTCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGAGGCCAGGGCAG
 ^
 303 ALWN1,

 ThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPhe
 362 ACTGTCCCCATCCCAACATCGAGGAGGTTGCTCTGTCACCACCGGAGAGATCCCTTT
 TGACACGGGGTAGGGTTGTAGCTCTCCAACGAGACAGGTGGTGGCCTCTAGGGAAAA

 TyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHis
 422 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGAGACATCTCATCTGTCAT
 ATGCCGTTCCGATAGGGGAGCTTCACTAGTTCCCCCTCTGAGTAGAGTAAAGACAGTA

 SerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaVal
 482 TCAAAGAAGAAGTGCACGAACTCGCCGAAAGCTGGTCGCATTGGGCATCAATGCCGTG
 AGTTCTTCTCACGCTGCTTGACGGCGTTCGACCGAGCGTAACCGTAGTTACGCCAC

 AlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValValVal
 542 GCCTACTACCGGGCTTGACGTGTCGTCATCCCGACCAGCGCGATGTTGTCGTGCG
 CGGATGATGGGCCAGAACCTGCACAGGCAGTAGGGCTGGTCGCCGCTAACACAGCAC
 ^ .
 550 SAC2, 560 DRD1,

 AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsn
 602 GCAACCGATGCCCTCATGACCGGCTATACCGGCACTTCGACTCGGTGATAGACTGCAAT
 CGTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACGTTA
 ^

FIGURE 18 - Page 2

662 ThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIle
 ACGTGTGTCACCCAGACAGTCGATTCAGCCTGACCCCTACCTCACCATGGAGACAATC
 TGCACACAGTGGGTCTGTCAGCTAAAGTCGGAATGGGATGGAAGTGGTAACTCTGTTAG

722 ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgArgThrGlyArgGlyLys
 ACGCTCCCCAAGATGCTGCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAAG
 TGCAGGGGGTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCCTGACCCTCCCCCTTC

782 ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 CCAGGCATCTACAGATTGTGGCACCGGGGAGCCTCCGGCATGTTGACTCGTCC
 GGTCCGTAGATGTCTAACACCGTGGCCCCCTCGCGGGAGGCCGTACAAGCTGAGCAGG
 816 BGLI, 833 DRD1,

842 ValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThr
 GTCCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGTGAGCTCACGCCCGAGACT
 CAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAGTGCAGGGCGCTCTGA
 881 SACI,

902 ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 ACAGTTAGGCTACGAGCGTACATGAACACCCGGGCTTCCGTGTGCCAGGACCATCTT
 TGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGCTGGTAGAA

931 SMAI XMAI,

962 GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln
 GAATTTGGGAGGGCGTCTTACAGGCTCACTCATATAGATGCCACTTCTATCCCAG
 CTTAAAACCCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTC
 985 STUI,

1022 ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
 ACAAAAGCAGAGTGGGAGAACCTTCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
 TGTTCGTCTCACCCCTCTTGGAGGAATGGACCATCGCATGGTCGGTGGCACACGCGA
 1069 DRA3,

1082 ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLys
 AGGGCTCAAGCCCTCCCCATCGTGGGACCAGATGTGGAAAGTGTGTTGATTGCCTCAAG
 TCCCAGTTGGGGAGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTC

1142 ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 CCCACCCCTCCATGGGCCAACACCCCTGCTATACAGACTGGCGCTGTTGAGAATGAAATC
 GGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCCGCGACAAGTCTTACTTAG
 1150 NCOI,

1202 ThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluVal
 ACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCGACCTGGAGGT
 TGGGACTGCGTGGGTCACTGGTTATGTAGTACTGTACGTACAGCCGGCTGGACCTCCAG
 1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,

1262 ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 GTCACGAGCACCTGGGTGCTCGTTGGCGCGTCCGGCTGCTTGGCCCGTATTGCCTG

FIGURE 18 - Page 3

CAGTGCTCGTGGACCCACGAGCAACGCCGCAGGACCGACGAAACGGCGATAACGGAC

SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIle
 1322 TCAACAGGCTCGTGGTCATAGTGGGCAGGGTCGTCTGTCCGGGAAGCCGGCAATCTA
 AGTTGTCGACGCACCAGTATCACCCGTCAGCAGAACAGGCCCTCGGCCGTTAGTAT

1369 NAEI,

ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
 1382 CCTGACAGGGAAAGTCTCTACCGAGAGTCGATGAGATGGAAGAGTGCTCTCAGCACTTA
 GGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTCTCACGAGAGTCGTGAAT

1385 DRD1,

ProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAACGGCCCTCGGCCTC
 GGATGTAGCTCGTCCCTACTACGAGCGGCTCGTCAAGTTCGTCAGTCCGGGAGCCGGAG

LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCGTCCCGTCAGGCAGAGTTATGCCCTGCTGTCCAGACCAACTGGCAA
 GACGTCGCGCAGGGCAGTCCGTCCTCAATAGCGGGGACGACAGGTCTGGTGACCGTT

1502 PSTI, 1507 TTH3I,

LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
 1562 AAACTCGAGACCTTCTGGCGAAGCATATGTTGAACTTCATCAGTGGATACAATACTTG
 TTTGAGCTCTGAAAGACCCGCTCGTATACACCTGAAGTAGTCACCCATGTTATGAAC

1565 XHOI, 1586 NDEI,

AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla
 1622 CGGGCTTGTCAACGCTGCCATTGCTTCAACATATTGGCTTTACAGCT
 CGCCCGAACAGTTGCGACGGACCATTGGGGCGTAACGAAGTAACCGAAAATGTCCA

1643 BSTE2, 1677 ALWN1 PVU2,

AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp
 1682 GCTGTCACCAGCCCCTAAACCACTAGCCAAACCCCTCCTCTTCAACATATTGGGGGGGTGG
 CGACAGTGGTCGGGTGATTGGTATCGGTTGGGAGGAGAAGTGTATAACCCCCCCCACC

ValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAla
 1742 GTGGCTGCCAGCTCGCCGCCGGTGCCTACTGCCTTGTGGCGCTGGCTTAGCT
 CACCGACGGGTGAGCGGGGGCCACGGCGATGACGGAAACACCGCGACCGAATCGA

1794 ESP1,

GlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyr
 1802 GGCGCCGCATCGGCAGTGTGGACTGGGAAGGTCTCATAGACATCCTGCAGGGTAT
 CCGCGCGGTAGCCGTACAACCTGACCCCTCAGGAGTATCTGTAGGAACGTCCCATA

1802 KAS1 NARI,

GlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSer
 1862 GGCAGGGCGTGGCGGGAGCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCCTCC
 CCGCGCCCGACCGCCCTCGAGAACACCGTAAGTCTAGTACTCGCCACTCCAGGGAGG

1878 SAC1, 1899 BSPH1,

FIGURE 18 - Page 4

1922 ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly
 ACGGAGGACCTGGTCAATCTACTGCCGCATCCTCTGCCCGAGCCCTCGTAGTCGGC
 TGCCTCCTGGACCAGTTAGATGACGGCGGTAGGAGAGCAGGGCCTCGGAGCATCAGCCG
 1928 TTH3I,
 ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
 1982 GTGGCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCAGGGGGCAGTGCAGTGG
 CACCAGACACGTGTTATGACGCCGGCGTGCAACCAGGGCCGCTCCCCCGTCACGTCACC
 2004 NAEI, 2017 SMAI XMAI,
 MetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
 2042 ATGAACCGGCTGTAGCCTCGCCTCCGGGGAACCATGTTCCCCCACGCACACTACGTG
 TACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAAGGGGTGCGTGTGAC
 2067 SMAI XMAI, 2093 DRA3,
 ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
 2102 CCGGAGAGCGATGCAGCTGCCCGCTACTGCCATACTCAGCAGCCTCACTGTAACCCAG
 GGCCTCTCGCTACGTCGACGGCGCAGTGACGGTATGAGTCGTGGAGTGACATTGGGTC
 2115 PVU2, 2159 ALWN1,
 LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
 2162 CTCCCTGAGGCGACTGCACCACTGGATAAGCTCGGAGTGTACCATGCTCCGGTTCC
 GAGGACTCCGCTGACGTGGCACCTATTGAGCCTCACATGGTGAGGTACGAGGCCAAGG
 2164 MST2, 2220 ECON1,
 TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
 2222 TGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCAGCTTAAGACCTGGCTA
 ACCGATTCCCTGTAGACCCCTGACCTATACGCTCCACAACCTCGCTGAAATTCTGGACCGAT
 LysAlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyr
 2282 AAAGCTAAGCTATGCCACAGCTGCCCTGGATCCCTTGTGTCTGCCAGCGCGGGTAT
 TTTCGATTGAGTACGGTGTGACGGACCCCTAGGGGAAACACAGGACGGTCGCGCCATA
 2285 ESP1, 2300 PVU2, 2310 BAMHI,
 LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 2342 AAGGGGCTGGCGAGGGGACGGCATATGCACACTCGCTGCCACTGTGGAGCTGAGATC
 TTCCCCCAGACCGCTCCCTGCCGTAGTACGTGTGAGCAGGGTACACCTCGACTCTAG
 ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 2402 ACTGGACATGTCAAAACGGGACGATGAGGATCGTCGGCTTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTGCCCTGCTACTCCTAGCAGCCAGGATCTGGACGTCCCTGTAC
 2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PST1,
 TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla
 2462 TGGAGTGGGACCTTCCCCATTAATGCCACACCACGGGCCCCCTGTACCCCCCTTCCTGCG
 ACCTCACCCCTGGAAGGGTAATTACGGATGTGGTGCCGGGACATGGGGGAAGGACGC
 2480 ASE1, 2497 APA1,

FIGURE 18 - Page 5

ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln
 2522 CCGAACTACACGTCTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCAG
 GGCTTGATGTGCAAGCGCGATACCTCCACAGACGTCTCCTTATGCACCTCTATTCCGTC

2553 PST1,

ValGlyAspPheHisTyrValThrGlyMetThrAspAsnLeuLysCysProCysGln
 2582 GTGGGGGACTTCACTACGTGACGGGTATGACTACTGACAATCTAAATGCCGTGCCAG
 CACCCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTACGGGCACGGTC

2594 DRA3,

ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro
 2642 GTCCCCATCGCCCGAATTTTACAGAATTGGACGGGTGCGCCTACATAGGTTGCGCCC
 CAGGGTAGCGGGCTTAAAAAGTGTCTAACCTGCCACGCGGATGTATCCAAACGCGGG

2702 ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTGCTGCGGGAGGAGGTATCATTAGACTCCACGAATACCCG
 GGGACGTTGGAAACGACGCCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGC

2757 HGIE2,

ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
 2762 TAGGGTCGCAATTACCTTGGAGGCCGAACCGGACGTGGCGTGTGACGTCCATGCTC
 CATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAGGTACGAG

2809 AAT2,

ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
 2822 ACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGCGAAGGTGGCGAGGGGATCACCC
 TGACTAGGGAGGGTATATTGTCGTCTCGCCGGCCCGCTTCCAACCGCTCCCTAGTGGG

2850 EAG1 XMA3,

ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
 2882 CCCTCTGTGGCCAGCTCTCGGCTAGCCAGCTATCCGCTCCATCTCAAGGCAACTTGC
 GGGAGACACCGGTGAGGAGCGATCGGTGATAGGGAGGTAGAGAGTTCCGTTGAACG

2889 BAL1, 2903 NHEI,

ThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGln
 2942 ACCGCTAACCATGACTCCCTGATGCTGAGCTCATAGAGGCCAACCTCTATGGAGGCAG
 TGGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCCGTC

2966 ESP1, 2969 SAC1,

GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 3002 GAGATGGCGGCCAACATCACCAAGGGTTGAGTCAGAAAAACAAAGTGGTATTCTGGACTCC
 CTCTACCCGCCGTTGAGTGGCTCCAACTCAGTCTTGTTCACCACAAAGACCTGAGG

3062 PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 TTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTG
 AAGCTAGGCGAACACCGCCCTCCCTGCTCGCCCTAGAGGCATGGCGTCTTAGGAC

3096 BGL2,

ArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro

FIGURE 18 - Page 6

3122 CGGAAGTCTGGAGATTGCCAGGCCCTGCCGTTGGCGGCCGGACTATAACCC
 GCCTTCAGAGCCTAAGCGGGTCCGGACGGCAAACCCGCACGGCTGATATTGGGG
 ^
 3143 ALWN1, 3164 EAG1 XMA3,
 .
 ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCGCTAGTGGAGACGTGGAAAAAGCCCAGTACGAACCACCTGTGGTCCATGGCTGCCCG
 GGCGATCACCTCTGCACCTTTGGCTGATGCTGGTGACACCAGGTACCGACGGC
 ^
 3217 HGIE2, 3229 NCOI,
 .
 LeuProProProLysSerProProValProProArgLysLysArgThrValValLeu
 3242 CTTCCACCTCAAAGTCCCCTCCTGTGCCTCGCCTCGGAAGAACGGGACGGTGGTCC
 GAAGGTGGAGGTTCAAGGGAGGACACGGAGGCGGAGCCTCTTCGCCTGCCACCAGGAG
 ^
 ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer
 3302 ACTGAATCAACCCATCTACTGCCTGGCGAGCTGCCACCAGAACGCTTGGCAGCTCC
 TGACTTAGTTGGGATAGATGACGGAACCGGCTGAGCGGTGGCTTCGAAACCCTCGAGG
 ^
 3332 SACI, 3346 HIND3,
 .
 SerThrSerGlyIleThrGlyAspAsnThrThrSerSerGluProAlaProSerGly
 3362 TCAACTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCGCCCTCTGGC
 AGTTGAAGGCCGTAATGCCGCTGTTATGCTGTTAGGAGACTCGGGCGGGAAAGACCG
 ^
 CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluPro
 3422 TGCCCCCCCAGCTCCGACGCTGAGTCCTATTCCCTCATGCCGCCCTGGAGGGGGAGCCT
 ACGGGGGGGCTGAGGCTGCAGGATAAGGAGGTACGGGGGGACCTCCCCCTCGGA
 ^
 3437 EAM11051,
 .
 GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTCAAGGTCAACGGTCAGTAGTGAGGCCAACGGAG
 CCCCTAGGCCAGAATCGCTGCCAGTACCGTCCAGTCATCACTCCGGTTGCGCCTC
 ^ ^
 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
 .
 AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla
 3542 GATGCGTGTGCTGCTCAATGTCTACTCTGGACAGGCGCACTCGTCACCCGTGCGCC
 CTACAGCACACGAGTTACAGAATGAGAACCTGCGCGTGAGCAGTGGGGCACGCGG
 ^
 3589 DRA3, 3600 SAC2,
 .
 AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
 3602 GCGGAAGAACAGAAACTGCCCATCAATGCACTAACGAAACTCGTGTACGTACCCACAAAT
 CGCCTTCTTGTCTTGACGGTAGTTACGTGATTGAGCAACGATGCACTGGGTGTTA
 ^
 3611 ALWN1, 3655 PFLM1,
 .
 LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
 3662 TTGGTGTATTCCACCCACCTCACGGAGTGGCTTGCGTCACGAACGGTTCCGCTCTTCAGTGTAAACTG
 AACACACATAAGGTGGTAGCTGGACTCGTCACGAACGGTTCCGCTCTTCAGTGTAAACTG
 ^
 3681 DRA3,
 .
 ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
 3722 AGACTGCAAGTTCTGGACAGCCATTACCAAGGACGTACTCAAGGAGGTAAAGCAGCGCG
 ^
 .

FIGURE 18 - Page 7

TCTGACGTTCAAGACCTGTCGGTAATGGCCTGCATGAGTTCCAATTCGTGCCGC
 3782 SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
 TCAAAAGTGAAGGCTAACCTGCTATCCGTAGAGGAAGCTGAGCCTGACGCCACAC
 AGTTTCACCTCCGATTGAACGATAGGCATCTCGAACGTCGGACTGCGGGGTGTG
 3816 HIND3,
 SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
 3842 TCAGCCAAATCCAAGTTGGTATGGGCAAAAGACGTCCGTTGCCATGCCAGAAAGGCC
 AGTCGGTTAGGTTCAAACCAATACCCGTTCTGAGGCAACGGTACGGTCTTCCGG
 3875 AAT2, 3890 BGLI,
 ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp
 3902 GTAACCCACATCAACTCCGTGGAAAGACCTCTGGAAGACAATGTAACACCAATAGAC
 CATTGGGTGTAGTTGAGGCACACCTTCTGGAAGACCTCTGTTACATTGTGGTTATCTG
 ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArgLys
 3962 ACTACCACATGGCTAAGAACGAGGTTCTGCGTCAGCCTGAGAAGGGGGTCGTAAG
 TGATGGTAGTACCGATTCTGCTCCAAAGACGCAAGTCGGACTCTCCCCCAGCATTC
 ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
 4022 CCAGCTCGTCTCATCGTGTCCCCGATCTGGCGTGCCTGCGAAAGATGGCTTG
 GGTCCAGCAGAGTAGCACAAGGGCTAGACCGCACCGCACACGTTTCTACCGAAAC
 TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
 4082 TACGACGTGGTTACAAAGCTCCCTTGGCGTGATGGGAAGCTCTACGGATTCCAATAC
 ATGCTGCACCAATGTTCGAGGGGAAACGGGACTACCCCTCGAGGATGCCTAAGGTTATG
 SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
 4142 TCACCAGGACAGGGTTGAATTCTCGCAAGCGTGGAAAGCTCAAGAAAACCCCAATG
 AGTGGTCCTGCGCCAACTTAAGGAGCACGTTGCACCTTCAGGTTTTGGGTTAC
 4160 ECORI,
 GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
 4202 GGGTCTCGTATGATAACCCGCTGTTGACTCCACAGTCAGTGAGAGCGACATCCGTACG
 CCCAAGAGCATACTATGGCGACGAAACTGAGGTGTCAGTCTCGCTGTAGGCATGC
 4229 DRD1, 4236 ALWN1,
 GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
 4262 GAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGTGGCCATCAAGTCC
 CTCCTCGTTAGATGGTACAACACTGGAGCTGGGGTTGGCGCACCGGTAGTTCAAGG
 4301 BGLI, 4308 BALI,
 LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
 4322 CTCACCGAGGGCTTATGTTGGGGCCCTCTTACCAATTCAAGGGGGAGAACTGCGGC
 GAGTGGCTCTCGAAATACAACCCCGGGAGAATGGTTAAGTCCCCCTCTTGACGCCG
 4345 APAI,
 TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
 4382 TATCGCAGGTGCCCGCGAGCGCGTACTGACAACTAGCTGTGGTAACACCCCTCACTTGC
 ATAGCGTCCACGGCGCGCTCGCCGATGACTGTTGATCGACACCATTGTGGAGTGAACG

FIGURE 18 - Page 8

4442 TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
 TACATCAAGGCCGGCGAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTG
 ATGTAGTTCCGGGCCGTGGACAGCTCGCGTCCCAGGGTCTGACGTGGTACGAGCAC
 ^
 4452 SMA1 XMA1,
 CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
 4502 TGTGGCGACGACTTAGTCGTTATCTGAAAGCGCGGGGGTCCAGGAGGACGCAGCGAGC
 ACACCGCTGCTGAATCAGCAATAGACACTTCGCGCCCCCAGGTCTCCTGCGCCGCTCG
 ^ ^
 4508 DRD1, 4511 TTH3I,
 LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
 4562 CTGAGAGCCTCACGGAGGCTATGACCAGGTACTCCGCCCCCTGGGACCCCCAACAA
 GACTCTCGGAAGTGCCTCCGATACTGGTCATGAGGCAGGGGGGACCCCTGGGGGTGTT
 ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 4622 CCAGAATACGACTTGGAGCTCATAACATCATGCTCTCCAACGTGTCAGTCGCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTG
 ^
 4637 SAC1,
 GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
 4682 GGCCTGGAAAGAGGGTCTACTACCTCACCGTGACCCCTACAACCCCCCTCGCGAGAGCT
 CCGCGACCTTCTCCCAGATGATGGAGTGGCACTGGATGTTGGGGAGCGCTCTCGA
 ^
 4731 NRUI,
 AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 4742 GCGTGGGAGACAGCAAGACACACTCCAGTCATTCTGGCTAGGCAACATAATCATGTT
 CGCACCCCTCTGCTGTTCTGAGTACGTTAAGGACCGATCCGTTGATTAGTACAAA
 AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 4802 GCCCCCACACTGTGGCGAGGATGATACTGATGACCCATTCTTAGCGTCCTTATAGCC
 CGGGGGTGTGACACCCGCTCTACTATGACTACTGGTAAAGAAATCGCAGGAATATCGG
 ^
 4806 PFLM1, 4807 DRA3,
 ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu
 4862 AGGGACCAAGCTTGAACAGGCCCTGATTGCGAGATCTACGGGCTGCTACTCCATAGAA
 TCCCTGGTCGAACCTGTCCGGAGCTAACGCTCTAGATGCCCGACGATGAGGTATCTT
 ^
 4893 BGL2,
 ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
 4922 CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTCACTCCAC
 GGTGACCTAGATGGAGGTTAGTAAGTTCTGAGGTACCGGAGTCGGTAAAGTGAGGTG
 ^
 4954 NCOI,
 SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
 4982 AGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGATGCCTCAGAAAACCTGGGGTACCG
 TCAATGAGAGGTCCACTTAGTTATCCACCGCGTACGGAGTCTTGTAAACCCATGGC
 ^
 5015 SPHI, 5035 KPN1,

FIGURE 18 - Page 9

5042 ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly
 CCCTTGCAGCTGGAGACACCAGGGCCCGAGCGTCGCGCTAGGCTCTGCCAGAGGA
 GGGAAACGCTCGAACCTCTGTGGCCCGGGCTCGCAGGCGATCCGAAGACCAGGTCTCCT
 ^ ^
 5064 APAI, 5091 BALI,
 GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
 5102 GGCAGGGCTGCCATATGTGCAAGTACCTCTCAACTGGGCAGTAAGAACAAAGCTCAA
 CCGTCCCAGCGTATAACACCGTTATGGAGAAGTTGACCCGTATTCTGTTCGAGTT
 ^
 5113 NDEI,
 LeuThrProIleAlaAlaAlaGlyGlnLeuAspLeuSerGlyTrpPheThrAlaGlyTyr
 5162 CTCACTCCAATAGCGGGCCGCTGGCCAGCTGGACTTGTCCGGCTGGTCACGGCTGGCTAC
 GAGTGAGGTTATCGCCGGCGACCGGTGACCTGAACAGGCCGACCAAGTGCCGACCGATG
 ^ ^ ^
 5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,
 SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
 5222 AGCGGGGGAGACATTATCACAGCGTGTCTCATGCCCGCCCGCTGGATCTGGTTTGC
 TCGCCCCCTCTGTAAATAGTGTGACAGAGTACGGGCCGGGCGACCTAGACCAAAACG
 ^
 5240 DRA3,
 LeuLeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgMetSerThrAsn
 5282 CTACTCTGCTTGCTGCAGGGTAGGCATCTACCTCCTCCCCAACGAAATGAGCACGAAT
 GATGAGGACGAACGACGTCCCCATCCGTAGATGGAGGAGGGTTGGCTTACTCGTGTCTA
 ^
 5295 PSTI,
 ProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGlnAspValLysPhe
 5342 CCTAAACCTCAAAGAAAGACAAACGTAACACCAACCGCGGCCGAGGACGTCAAGTTC
 GGATTGGAGTTCTTCTGGTTGCATTGTGGTTGGCCGCCGGCTCTGCAGTTCAAG
 ^ ^ ^
 5380 NOTI, 5381 EAG1 XMA3, 5390 AAT2, 5401 SMA1 XMA1,
 ProGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeu
 5402 CCGGGTGGCGGTCAAGATCGTGGAGTTACTTGTGCGCCGCAGGGCCCTAGATTG
 GGCCCACCGCCAGTCAAGAACACCTCAAATGAACAAACGGCGCGTCCCCGGGATCTAAC
 ^
 5449 APAI,
 GlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnPro
 5462 GGTGTGCGCGCAGAGAAAAGACTTCCGAGCGGTGCAACCTCGAGGTAGACGTCAGCCT
 CCACACGCGCGCTGCTTTCTGAAGGCTGCCAGCGTTGGAGCTCCATCTGCAGTCGGA
 ^ ^ ^ ^
 5467 BSSH2, 5478 XMNI, 5502 XHOI, 5511 AAT2,
 IleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpPro
 5522 ATCCCCAAGGCTCGTGGCCCGAGGGCAGGACCTGGCTCAGCCGGTACCCCTGGCCC
 TAGGGGTTCCGAGCAGCCGGCTCCCGTCTGGACCCGAGTCGGGCCATGGGAACCGGG
 ^ ^ ^ ^
 5548 ALWN1, 5558 ESP1, 5564 SMA1 XMA1, 5568 KPNI,
 LeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArg
 5582 CTCTATGGCAATGAGGGCTGCCGGATGGCTCCTGTCTCCCGTGGCTCTCGG
 GAGATACCGTTACTCCGACGCCACCCGCCCTACCGAGGACAGAGGGCACCGAGAGCC

FIGURE 18 - Page 10

ProSerTrpGlyProThrAspProArgArgSerArgAsnLeuGlyLysValIleAsp
5642 CCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTGCGCAATTGGGTAAGGTCAICGAT
GGATCGACCCCGGGGTGTCTGGGGGCCCATCCAGCGCTTAAACCCATTCCAGTAGCTA
5650 APAI, 5696 CLAI,
ThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyAlaProLeu
5702 ACCCTACGTGCGGCTTCGCCGACCTCATGGGTACATACCGCTCGTCGGCGCCCTCTT
TGGGAATGCACGCCGAAGCGGCTGGAGTACCCCATGTATGGCGAGCAGCCGGGGAGAA
5724 HGIE2, 5750 KASI NARI, 5756 ECON1,
GlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeuGluAspGlyValAsnTyr
5762 GGAGGCCTGCCAGGGCCCTGGCGCATGGCGTCCGGTTCTGGAAAGACGGCGTGAACTAT
CCTCCCGACGGTCCCGGGACCGCGTACCGCAGGCCAAGACCTCTGCCGCAC TGATA
5772 BSTXI, 5775 APAI,
AlaThrGlyAsnLeuProGlyCysSerOC AM
5822 GCAACAGGGAACCTTCCCTGGTTGCTCTTAATAGTCGAC
CGTTGCCCTTGGAGGACCAACGAGAATTATCAGCTG
5854 SALI,

FIGURE 19

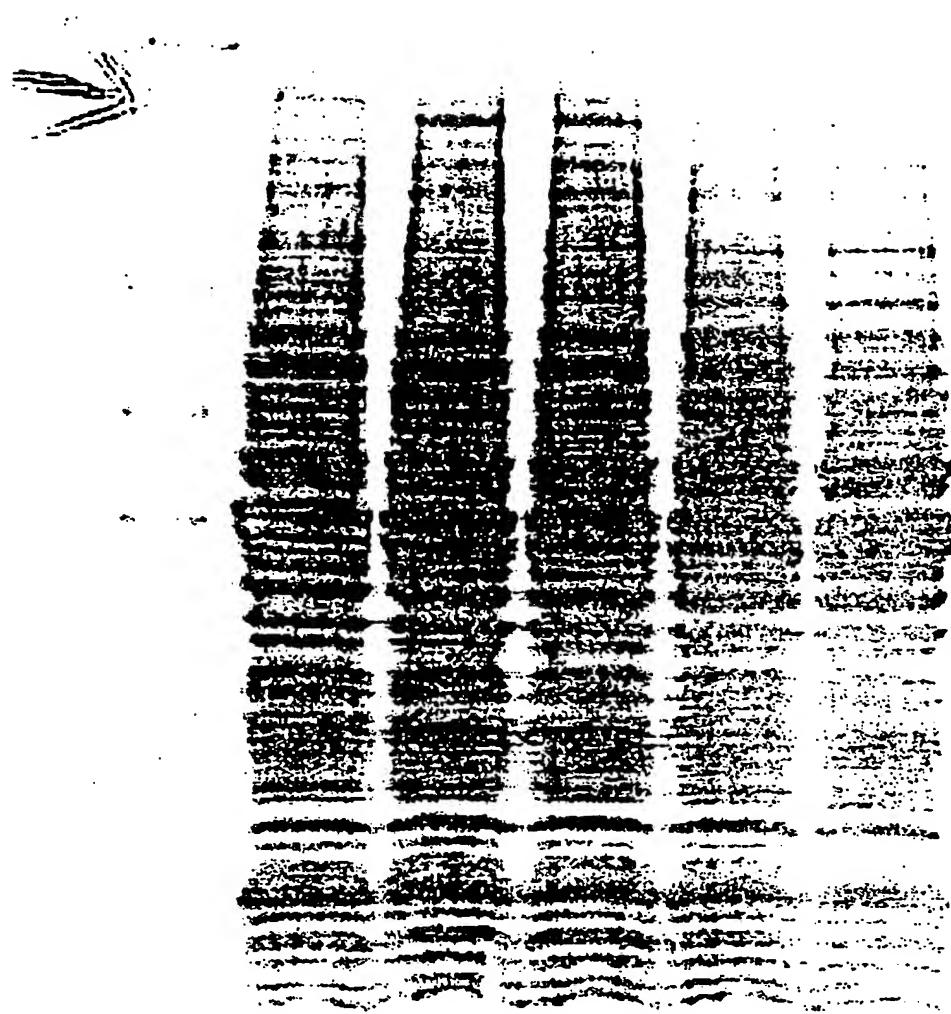


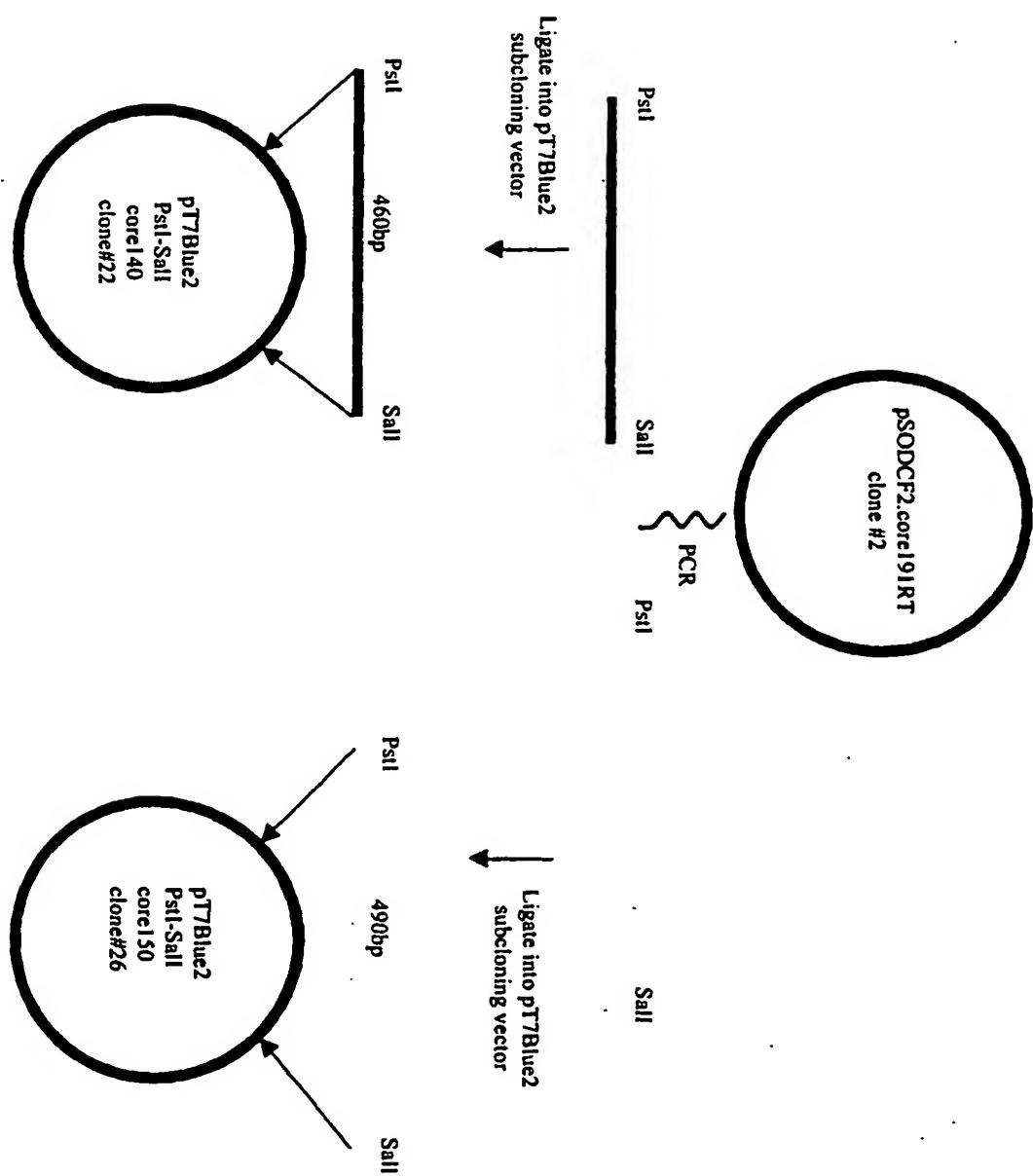
FIGURE 20 - Page 1

FIGURE 20 - Page 2

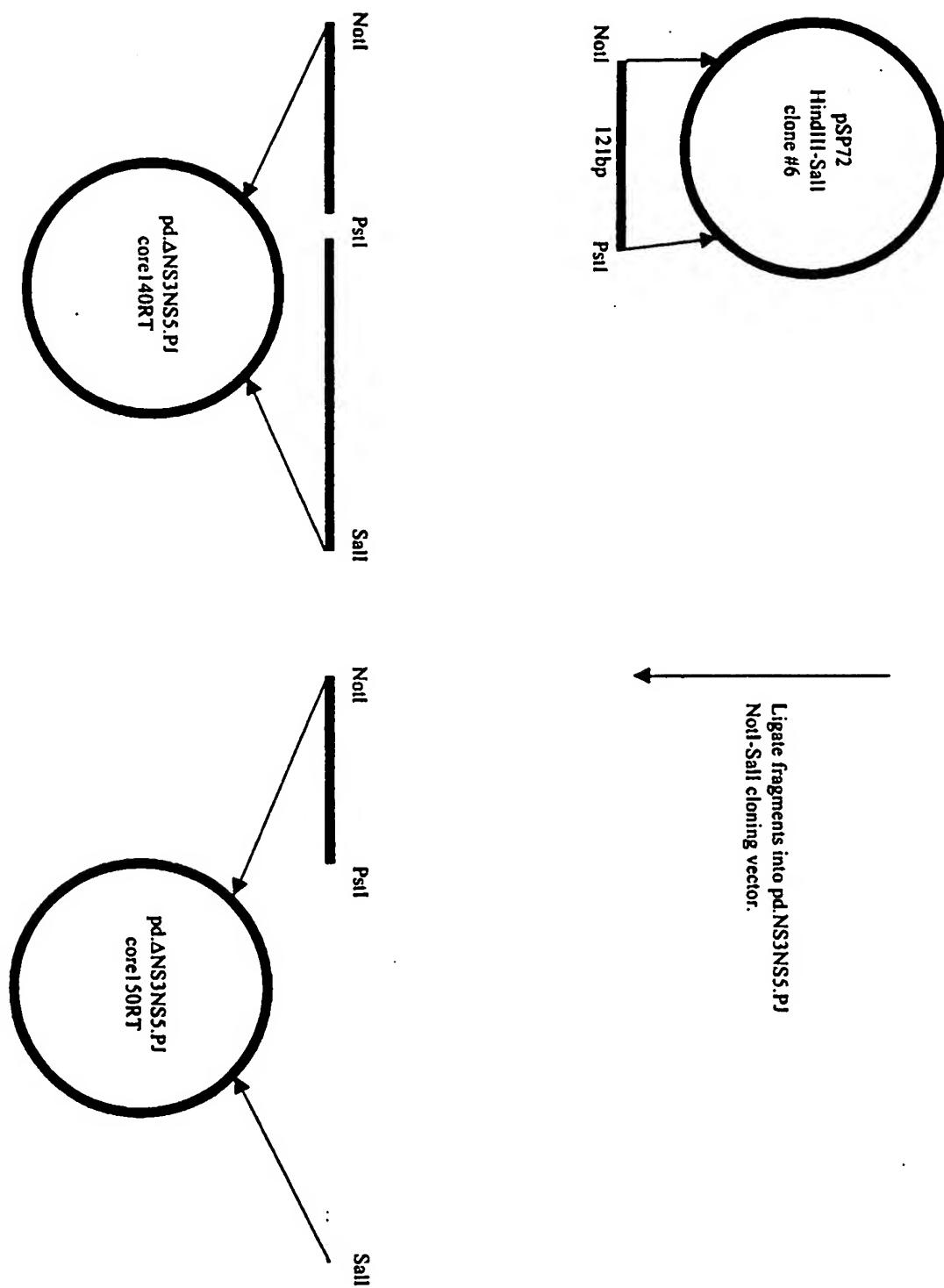


FIGURE 21 - Page 1

MetAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsn
 2 AGCTTACAAAACAAAATGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAAC
 TCGAATGTTTGTGTTACCGACGTATACTCGAGTCCCAGTATTCCACGATCATGAGTTG
 ^
 1 HIND3, 24 NDEI, 52 SCAI,

 ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAsp
 62 CCCTCTGTTGCTGCAACACTGGGCTTGCTTACATGTCCAAGGCTCATGGGATCGAT
 GGGAGACAAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTAGCCTAGCTA
 ^
 116 CLAI,

 ProAsnIleArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSerThr
 122 CCTAACATCAGGACCAGGGGTGAGAACAAATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTTGTAAATGGTGACCGTCGGGGTAGTGCATGAGGTGG

 TyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCys
 182 TACGGCAAGTCCCTGCGACGGCGGGTGCCTCGGGGGCGTTATGACATAATAATTGT
 ATGCCGTTCAAGGAACGGCTGCCGCCACGAGCCCCCGCGAATACTGTATTATTAAACA

 AspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGln
 242 GACGAGTGCCTCCACGGATGCCACATCCATCTGGCATTGGCAGTGTCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTAGGTAGAACCGTAACCGTGACAGGAAGTGGTT

 AlaGluThrAlaGlyAlaArgLeuValLeuAlaThrAlaThrProProGlySerVal
 302 GCAGAGACTGCGGGGGCGAGACTGGTTGTGCTGCCACCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCCCGCTCTGACCAACAGGAGCGGTGGCGGTGGGGAGGCCAGGGCAG
 ^
 303 ALWN1,

 ThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPhe
 362 ACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGCTCCACCACCGAGAGATCCCTTT
 TGACACGGGTAGGGTTGTAGCTCCTCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAA

 TyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHis
 422 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTGTGTCAT
 ATGCCGTTCCGATAGGGGGAGCTTCATTAGTCCCCCTCTGTAGAGTAGAACAGTA

FIGURE 21 - Page 2

482 SerLysLysLysCysAspGluL^{eu}AlaAlaLysLeuValAlaLeuGlyIleAsnAlaVa:
 TCAAAGAAGAAGTGGACGA^{ACTCGCCGAAAGCTGGTCGATTGGCATCAATGCCGTG}
 AGTTCTTCTTCACGCTGCTTGAGCGGCC^{TTCGACCAGCGTAACCGTAGTTACGGCAC}

 542 AlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValVal
 GCCTACTACCGCGGTCTTGACGTGTC^{CGTACATCCCACCAGCGGATGTTGTCGTG}
 CGGATGATGGCGCC^{AAGTGCACAGGAGTAGGGCTGGTCGCC}GTACAACAGCAC

 550 SAC2, 560 DRD1,

 AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsn
 602 GCAACCGATGCCCTCATGACCGCTATA^{CCGGCAGTCGACTCGGTGATAGACTGCAAT}
 CGTTGGCTACGGAGTACTGCC^{GATATGGCGCTGAAGCTGAGCCACTATCTGACGTTA}

 615 BSPH1,

 ThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIle
 662 ACGTGTGTCACCCAGACAGTCGATT^{TCAGCCTGACCCATTGAGACAATC}
 TGCACACAGTGGCTGTCAGCTAAAGTCGGA^{ACTGGGATGGAAGTGGTAAC}CTGTTAG

 722 ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 ACGCTCCCCCAAGATGCTGTC^{CCCGACTCAACGTCGGGGCAGGACTGGCAGGGGAAG}
 TGCGAGGGGGTTCTACGACAGAGGGCGT^{GAGTTGCAGCCCCGTCTGACCGTCCCC}CTTC

 782 ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 CCAGGCATCTACAGATTG^{TGGCACCGGGGAGCGCCCTCCGCATGTTGACTCGTCC}
 GTC^{CGTAGATGTC}AAACACCGTGGCCCC^{TGCGGGGAGGCC}GTACAAGCTGAGCAGG

 816 BGLI, 833 DRD1,

 ValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThr
 842 GTCCTCTGTGAGT^{GCTATGACGCAGGCTGCTGGTATGAGCTCACGCCCCGAGACT}
 CAGGAGACACTCACGATACT^{TGCGTCCGACACGAACCATACTCGA}GTGCGGGGCTCTGA

 881 SAC1,

 ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 902 ACAGTTAGGCTACGAGCGTACATGAACACCCGGGCT^{CCCCTGGCAGGACCATCT}
 TGTCAATCCGATGCTCGCATGTACTTG^{GGGGCCCCGAAGGGCACACGGT}CCTGGTAGAA

 931 SMA1 XMA1,

 GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln
 962 GAATTG^{GGGAGGGCGTCTTACAGGCCTCACTCATATAGATGCCACTTCTATCCCAG}
 CTTAAAACCC^{TCCCGAGAAATGTCGGAGTGAGTATATCTACGGGT}GAAAGATAGGGTC

 985 STUI,

 ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
 1022 ACAAAAGCAGAGTGGGGAGAAC^{CTTCCTACCTGGTAGCGTACCAAGCCACCGT}
 TGTTCTGCTCACCC^{CTTGGAAAGGAATGGACCATCGCATGGT}GGCACACGC^{GA}

 1069 DRA3,

 ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLys
 1082 AGGGCTCAAGCCCTCCCCC^{CATCGTGGGACCAGATGTGAAAGTGT}TTGATTGCC^{CTCAAG}

FIGURE 21 - Page 3

TCCCGAGTTGGGGAGGGGGTAGCACCTGGTACACCTTCACAAACTAACGGAGTT
 1142 ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 CCCACCCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGGCCTGTTCAAATGAAATC
 GGGTGGGAGGTACCCGGTGTGGGACGATATGTCTGACCCGCACAAGTCTTACTTAG
 1150 NCOI,
 ^
 ThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluVal
 1202 ACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTC
 TGGGACTGCGTGGGTCACTGGTTATGTAGTACTGTACGTACAGCCGGCTGGACCTCCAG
 ^ ^ ^ ^
 1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,
 ^
 ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 1262 GTACAGACACCTGGGTGCTCGTGGCGCGTCCGGCTGGCTGGCCGCGTATTGCCTG
 CAGTGCTCGTGGACCCAGCAGCAACCGCCGCAGGACCGACGAAACCGGCCATAACGGAC
 ^
 SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIle
 1322 TCAACAGGCTGCGTGGTCATAGTGGCAGGGTGTCTGTCGGGAAGCCGGCAATCATA
 AGTTGTCCGACCGACCAGTATCACCCGCTCCAGCAGAACAGGCCCTCGGCCGTTAGTAT
 ^
 1369 NAEI,
 ^
 ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
 1382 CCTGACAGGGAAAGTCCTCTACCGAGAGTTGATGAGATGGAAGAGTGCTCTCAGCACTTA
 GGACTGTCCCTTCAGGAGATGGCTCTAACGCTACTCACCTCTCACGAGAGTCGTGAAT
 ^
 1385 DRD1,
 ^
 ProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAACGGCCCTCGGCCCTC
 GGCATGTAGCTCGTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTCCGGAGCCGGAG
 ^
 LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCGTCCCGTCAGGCAGAGGTTATGCCCTGCTGTCCAGACCAACTGGCAA
 GACGTCTGGCGCAGGGCAGTCCGTCTCAAATAGCGGGGACGACAGGTCTGGTTGACCGTT
 ^
 1502 PSTI, 1507 TTH3I,
 ^
 LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
 1562 AACTCGAGACCTCTGGCGAAGCATATGTGAACTTCATCAGTGGATACAATACTTG
 TTTGAGCTCTGGAAGACCCGTTCGTATACACCTTGAAGTAGTCACCCATGTTATGAAC
 ^
 1565 XHOI, 1586 NDEI,
 ^
 AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla
 1622 GCGGGCTTGTCAACGCTGCCCTGGTAACCCGCCATTGCTCATTGATGGCTTTACAGCT
 CGCCCGAACAGTGCAGGGACATTGGGGCGTAACGAAGTAACCGAAAATGTCGA
 ^
 1643 BSTE2, 1677 ALWN1 PVU2,
 ^
 AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp
 1682 GCTGTCACCAGCCCCTAAACCACTAGCCAAACCCCTCCTTCAACATATGGGGGGGTGG
 CGACAGTGGTCGGGTGATTGGTATCGGTTGGGAGGAGAAGTTGTATAACCCCCCACC

FIGURE 21 - Page 4

1742 Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala
 GTGGCTGCCAGCTGCCGCCCCGGCTACTGCCTTGTGGCGCTGGCTTAGCT
 CACCGACGGGTCGAGCGGGGGCACGGCGATGACGAAACACCCGCGACCGAATCGA
 ^
 1794 ESP1,
 Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr
 1802 GGC GCC GCC AT CGG CAG T GT GG ACT GGG GAA AGG C CT CAT AG AC AT CCT TG CAG GG T AT
 CCG CGG CGG TAG CGT C ACA ACC TG ACC C TT CC AGG AGT AT CT GT TAG GA AC GT CCC ATA
 ^
 1802 KAS1 NARI,
 Gly Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser
 1862 GGC CGGG CGT GG CGG GAG CT CT GT GG ATT CA AG AT CAT GAG CG GT GAG GT CCC CT CC
 CCG CGC CC CG CAC CG CC CT CG AGA AAC ACC GT AA GT CT AG T ACT CG CC ACT CC CAG GGG AGG
 ^
 1878 SAC1, 1899 BSPH1,
 Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly
 1922 ACG GAG GAC CT GG T CA AT CT ACT G C C C G C AT C C T C G C C C G G A G C C C T C G T A G T C G G C
 T G C C T C C T G G A C C A G T T A G A T G A C G G G C G G T A G G A G A G C G G G C T C G G A G C A T C A G C C G
 ^
 1928 TTH3I,
 Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp
 1982 GTGGTCTGTG CAG CA AAT ACT G C C C G G C A C G T T G G C C C G G C A G G G G C A G T G C A G T G G
 C A C C A G A C A C G T C G T T A T G A C G C G G C C G T G C A A C C G G C C C G C T C C C C G T C A C G T C A C C
 ^
 2004 NAEI, 2017 SMAI XMAI,
 Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val
 2042 ATGAACC CGG CTG ATAG C C T C G C C T C C C G G G A A C C A T G T T C C C C C A C G C A C T A C G T G
 T A C T T G G C C G A C T A T C G G A A G C G G A G G G C C C C T T G G T A C A A A G G G G T G C G T G A T G C A C
 ^
 2067 SMAI XMAI, 2093 DRA3,
 Pro Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln
 2102 CCG GAG A G C G A T G C A G T G C C C G C G T C A C T G C C A T A C T C A G C A G C T C A C T G T A A C C A G
 G G C C T C T C G C T A C G T C G A C G G G C G C A G T G A C G G T A T G A G T C G T C G G A G T G A C A T T G G G T C
 ^
 2115 PVU2, 2159 ALWN1,
 Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser
 2162 C T C C T G A G G C G A C T G C A C C A G T G G A T A A G C T G G A G T G T A C C A C T C C A T G C T C C G G T T C C
 G A G G A C T C C C G T G A C G T G G T C A C C T A T T C G A G C C T C A C A T G G T G A G G T A C G A G G C C A A G G
 ^
 2164 MST2, 2220 ECON1,
 Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu
 2222 T G G C T A A G G G A C A T C T G G G A C T G G A T A T G C G A G G T G T G A G C G A C T T T A A G A C C T G G C T A
 A C C G A T T C C C T G T A G A C C C T G A C C T A T A C G C T C C A C A A C T C G C T G A A A T T C T G G A C C G A T
 ^
 2282 Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr
 AAAGCTAAGCTCATGCCACAGCTGCCGGATCCCCTTGTGTCCTGCCAGCGCGGGTAT
 T T C G A T T C G A G T A C G G T G C G A C G G A C C C T A G G G G A A A C A C A G G A C G G T C G C G C C C A T A
 ^
 2285 ESP1, 2300 PVU2, 2310 BAMHI,

FIGURE 21 - Page 5

2342 LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 AAGGGGGCTGGCGAGGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATC
 TTCCCCCAGACCGTCCCTGCCGTAGTAGTGTGAGCGACGGTGACACCTCGACTCTAG

2402 ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 ACTGGACATGTCAAAACGGGACGATGAGGATCGCTGGCTTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTGCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCTGTAC
 ^ ^ ^ ^

2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,

2462 TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla
 TGGAGTGGGACCTCCCCATTAAATGCCTACACCACGGGCCCTGTACCCCCCTTCCTGCG
 ACCTCACCCCTGGAAAGGGTAATTACGGATGTGGTGCCCAGGACATGGGGGAAGGACGC
 ^ ^

2480 ASE1, 2497 APAI,

2522 ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln
 CCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCAG
 GGCTTGATGTGCAAGCGCGATACTCCCACAGACGTCTCCTATGCACCTCTATTCCGTC
 ^

2553 PSTI,

2582 ValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGln
 GTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTTAAATGCCGTGCCAG
 CACCCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTACGGCACGGTC
 ^

2594 DRA3,

2642 ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro
 GTCCCACGCCCGAATTTCACAGAATTGGACGGGTGCGCCTACATAGGTTGCGCCC
 CAGGGTAGCGGGCTTAAAAAGTGTCTAACCTGCCCCACCGGATGTATCCAACGCCGG

2702 ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTTGCTCGGGAGGAGGTATCATTAGAGTAGGACTCCACGAATAACCG
 GGGACGTTGGAACGACGCCCTCCCATAGTAAGTCTCATCCTGAGGTGCTTATGGC
 ^

2757 HGIE2,

2762 ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
 GTAGGGTCGCAATTACCTTGCAGGCCGAACGGACGTGGCGTGTGACGTCCATGCTC
 CATCCCAGCGTTAATGAAACGCTCGGCTTGGCCTGCACCGGACAACGTGAGGTACGAG
 ^

2809 AAT2,

2822 ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
 ACTGATCCCTCCCATAAACAGCAGAGGCGGCCGAAGGTTGGCGAGGGATCACCC
 TGACTAGGGAGGGTATATTGTCGTCTCCGCCGCGCTTCCAACCGCTCCCTAGTGGG
 ^

2850 EAG1 XMA3,

2882 ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
 CCCTCTGTGCCAGCTCTCGGCTAGCCAGCTATCCGCTCATCTCAAGGCAAATTGC
 GGGAGACACCGGTCGAGGAGGCCATCGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACG
 ^

2889 BALI, 2903 NHEI,

FIGURE 21 - Page 6

2942 ThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGin
 ACCGCTAACCATGACTCCCTGATGCTGAGCTCATAGAGGCCAACCTCTATGGAGGCAG
 TGGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGGTTGGAGGATAACCTCCGTC
 ^ ^
 2966 ESP1, 2969 SACI,

 3002 GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 GAGATGGCGGCAACATCACCAAGGGTGAGTCAGAAAAACAAAGTGGTATTCTGGACTCC
 CTCTACCCGCCGTTGTAGTGGTCCCACTCAGTCTTTGTTCAACCCTAACGACCTGAGG

 3062 PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 TTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCTG
 AAGCTAGGCAGAACACCGCCTCCTGCTCGCCCTCTAGAGGCATGGCGTCTTAGGAC
 ^
 3096 BGL2,

 3122 ArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro
 CGGAAGTCTGGAGATTGCCAGGCCCTGCCGTTGGCGCGCCGACTATAACCCC
 GCCTTCAGAGCCTCTAACGGGGTCCGGGACGGCAAACCCGCGCCGCTGATATTGGGG
 ^
 3143 ALWN1, 3164 EAG1 XMA3,

 3182 ProLeuValGluThrTrpLysProAspTyrGluProProValValHisGlyCysPro
 CCGCTAGTGGAGACGTGGAAAAAGCCGACTACGAACCACCTGTGGTCCATGGCTGCCCG
 GGCATCACCTCTGACACCTTTCGGGCTGATGCTGGTGACACCAGGTACCGACGGGC
 ^ ^
 3217 HGIE2, 3229 NCOI,

 3242 LeuProProProLysSerProProValProProArgLysLysArgThrValValLeu
 CTTCACCTCAAAGTCCCTCTGTGCCTCCGCCTCGAAGAACGGACGGTGGTCTC
 GAAGGTGGAGGTTCAAGGGAGGACACGGAGGCGAGCCTTCGCCTGCCACCAGGAG

 3302 ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer
 ACTGAATCAACCCATCTACTGCCTGGCGAGCTCGCCACCAAGCTTGGCAGCTCC
 TGACTTAGTTGGATAGATGACGGAACCGGCTCGAGCGGTGGCTTCGAAACCGTCGAGG
 ^ ^
 3332 SACI, 3346 HIND3,

 3362 SerThrSerGlyIleThrGlyAspAsnThrThrSerSerGluProAlaProSerGly
 TCAACTCCGGCATTACGGCGACAATACGACAACATCCTCTGAGCCCCTCTGGC
 AGTGAAGGCCGTAATGCCGCTGTTATGCTGTTAGGAGACTCGGGGGAAAGACCG

 3422 CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluPro
 TGCCCCCCCCGACTCCGACGCTGAGTCTATTCCCTCATGCCCTGGAGGGGAGCCT
 ACGGGGGGGCTGAGGCTCGACTCAGGATAAGGAGGTACGGGGGGACCTCCCCCTCGGA
 ^
 3437 EAM11051,

 3482 GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 GGGGATCCGGATCTTAGCGACGGGCTATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAG
 CCCCTAGGCCTAGAATCGCTGCCAGTACCAAGTGGCCAGTCATCACTCCGGTTGCGCCTC
 ^ ^ ^
 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,

 3542 AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla
 GATGTCGTGTGCTGCTCAATGTCTTACTCTGGACAGGCGCAGTCACCCGTGCGCC
 CTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGCACGCGG

FIGURE 21 - Page 7

3589 DRA3, 3600 SAC2,

3602 AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
 GCGGAAGAACAGAAACTGCCATCAATGCACTAAGCAACTCGTGCTACGTACCACAAT
 CGCCTTCTTGCTTTGACGGTAGTTACGTGATTGAGCAACGATGCAGTGGTGTAA

3611 ALWN1, 3655 PFLM1,

3662 LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
 TTGGTGTATTCCACCACCTCACGCAGTGCTGCCAAAGCAGAAAGTCACATTGAC
 AACCACATAAGGTGGTAGTCACGAAACGGTTCCGTCTTCAGTGTAAACTG

3681 DRA3,

3722 ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
 AGACTGCAAGTCTGGACAGCCATTACCAAGGACGTACTCAAGGAGGTTAAAGCAGCGCG
 TCTGACGTTCAAGACCTGTCGGTAATGGCCTGCATGAGTCCCTCCAATTTCGTCGCCGC

3782 SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
 TCAAAAGTGAAGGCTAACCTGCTATCCGTAGAGGAAGCTTGACGCCCTGACGCCACAC
 AGTTTCACTTCCGATTGAACGATAGGCATCTCCTCGAACGTCGGACTGCGGGGTGTG

3816 HIND3,

3842 SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
 TCAGCCAAATCCAAGTTGGTTATGGGCAAAGACGTCCGTTGCCATGCCAGAAAGGCC
 AGTCGTTAGGTTCAAACCAATACCCGTTTCTGCAGGCAACGGTACGGTCTTCGG

3875 AAT2, 3890 BGLI,

3902 ValThrHisIleAsnSerValTrpLysAspLeuGluAspAsnValThrProIleAsp
 GTAACCCACATCAACTCCGTGGAAAGACCTCTGGAAAGACAATGTAACACCAATAGAC
 CATTGGGTGTAGTTGAGGCACACCTTCTGGAAAGACCTCTGTTACATTGTGGTTATCTG

3962 ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArgLys
 ACTACCATCATGGCTAACGAGGTTTCTGCGTCAGCCTGAGAAGGGGGTCGTAAG
 TGATGGTAGTACCGATTCTGCTCAAAGACGCAAGTCGGACTCTCCCCCAGCATTC

4022 ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
 CCAGCTCGTCTCATCGTGTCCCCGATCTGGCGTGCCTGCGAAAAGATGGTTTG
 GGTCGAGCAGAGTAGCACAAAGGGCTAGACCCGCACCGCACACGTTTCTACCGAAAC

4082 TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
 TACGACGTGGTTACAAAGCTCCCTTGGCGTGTGGAAAGCTCTACGGATTCCAATAC
 ATGCTGCACCAATGTTCGAGGGGACCGGCACACTCCCTCGAGGATGCCTAACGGTTATG

4142 SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
 TCACCAAGGACAGCGGGTTGAATTCTCGTCAAGCGTGGAAAGTCCAAGAAAACCCAATG
 AGTGGCCTGCGCCAACCTAACGGAGCACGTTCGCACCTCAGGTTCTTGGGTTAC

4160 ECORI,

4202 GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
 GGGTTCTCGTATGATAACCGCTGCTTGACTCCACAGTCACTGAGAGCGACATCCGTACG
 CCCAAGAGCATACTATGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGC

FIGURE 21 - Page 8

4229 DRD1, 4236 ALWN1,

4262 GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
 GAGGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCC
 CTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGTTGGGCCACCGTAGTTAGG

4301 BGLI, 4308 BALI,

4322 LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
 CTCACCGAGAGGCTTTATGTTGGGGCCCTTACCAATTCAAGGGGGAGAACTGCGGC
 GAGTGGCTCTCCGAAATACAACCCCGGGAGAATGGTTAAGTTCCCCCTCTTGACGCCG

4345 APAI,

4382 TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
 TATCGCAGGTGCCGCGCGAGCGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGC
 ATAGCGTCCACGGCGCGCTGCCGCATGACTGTTGATCGACACCATTGTGGAGTGAACG

4442 TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
 TAÇATCAAGGCCGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTG
 ATGTAGTTCCGGGCCGTCGGACAGCTGGCGTCCCAGGTCCTGACGTGGTACGAGCAC

4452 SMAI XMAI,

4502 CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
 TGTGGCGACGACTTAGTCGTATCTGTGAAAGCGGGGGTCCAGGAGGACGCCGGCAGC
 ACACCGCTGCTGAATCAGCAATAGACACTTCGCGCCCCCAGGTCCTCGCCGCTCG

4508 DRD1, 4511 TTH3I,

4562 LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
 CTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCTGGGACCCCCACAA
 GACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGGGTGTT

4622 ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 CCAGAACATACGACTTGGAGCTATAACATCATGCTCCTCCAACGTGTCAGTCGCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTG

4637 SACI,

4682 GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
 GGCCTGGAAAGAGGGTCTACTACCTCACCGTGACCCCTACAACCCCCCTCGCAGAGCT
 CGCGACCTTCTCCAGATGATGGAGTGGCACTGGATGTTGGGGAGCGCTCTCGA

4731 NRUI,

4742 AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 GCGTGGGAGACAGCAAGACACTCCAGTCATTCTGGCTAGGCAACATAATCATGTTT
 CGCACCCCTCTGTCGTTCTGTGAGGTCAGTTAAGGACCGATCCGTTGTTAGTACAAA

4802 AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 GCCCCCCACACTGTGGCGAGGATGATACTGATGACCCATTCTTAGCGTCTTATAGCC
 CGGGGGTGTGACACCCGCTCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCGG

4806 PFLM1, 4807 DRA3,

ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu

FIGURE 21 - Page

4862 AGGGACCAGCTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGAA
TCCCTGGTCGAACCTGTCCGGAGCTAACGCTCTAGATGCCCGACGATGAGGTATCTT

4893 BGL2,

4922 ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTCACTCCAC
GGTACGCTAGATGGAGGTTAGTAAGTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTG

4954 NCOI,

4982 SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
AGTTACTCTCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAATGGGTACCG
TCAATGAGAGGTCCACTTAGTTATCCCACCGGGTACGGAGTCTTTGAACCCATGGC

5015 SPHI, 5035 KPNI,

5042 ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly
CCCTTGCAGCTTGAGACACCAGGCCCCGGAGCGTCCCGCTAGGCTTCTGGCCAGAGGA
GGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCGCAGTCCGAAGACCAGGTCTCCT

5064 APAI, 5091 BALI,

5102 GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
GGCAGGGCTGCATATGTGGCAAGTACCTCTTCAACTGGGAGTAAGAACAAAGCTCAA
CCGCCCCACGGTATAACACCGTCATGGAGAAAGTTGACCCGTATTCTGTTGAGTT

5113 NDEI,

5162 LeuThrProIleAlaAlaAlaGlyGlnLeuAspLeuSerGlyTrpPheThrAlaGlyTyr
CTCACTCCAATAGCGGCCGCTGGCCAGCTGGACTTGTCCGGCTGGTTCACGGCTGGCTAC
GAGTGAGGTTATCGCCGGCGACCGGTGCGACTGAACAGGCCGACCAAGTGGCCACCGATG

5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,

5222 SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
AGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCGGCCCGCTGGATCTGGTTTGC
TCGCCCCCTGTAAATAGTGTGCGCACAGAGTACGGGCCGGGCGACCTAGACCAAAACG

5240 DRA3,

5282 LeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgMetSerThrAsn
CTACTCCTGCTGCTGCAGGGTAGGCATCTACCTCCTCCCCAACCGAATGAGCACGAAT
GATGAGGACGACGTCGGCATCCGTAGATGGAGGAGGGGTTGGCTTACTCGTGCTTA

5295 PSTI,

5342 ProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGlnAspValLysPhe
CCTAAACCTCAAAGAAAGACCAAACGTAACACCAACCGCGGGCGCAGGACGTCAAGTTC
GGATTTGGAGTTCTTCTGGTTGCATTGTGGTTGGCCGGCGTCTGCAGTTCAAG

5380 NOTI, 5381 EAG1 XMA3, 5390 AAT2, 5401 SMA1 XMA1,

5402 ProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeu
CCGGGTGGCGGTCAAGATCGTTGGAGTTACTTGTGCCCCGGCAGGGGCCCTAGATTG
GGCCCACCGCCAGTCTAGCAACCACCTCAAATGAACAAACGGCCGTCGGGATCTAAC

FIGURE 21 - Page 10

5449 APAI,
 5462 GlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGinPro
 GGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTGCAACCTCGAGGTAGACGTCAGCCT
 CCACACGCGCGCTGCTCTTCTGAAGGCTGCCAGCGTTGGAGCTCCATCTGCAGTCGGA
 , ^ ^ ^ ^
 5467 BSSH2, 5478 XMNI, 5502 XHOI, 5511 AAT2,
 5522 IleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpPro
 ATCCCCAAGGCTCGTCGGCCCAGGGCAGGACCTGGGCTCAGCCCAGGTACCCCTGGCCC
 TAGGGGTTCCGAGCAGCCGGCTCCGCTGGACCCGAGTCGGGCCATGGAACCGGG
 ^ ^ ^ ^
 5548 ALWN1, 5558 ESP1, 5564 SMAI XMAI, 5568 KPNI,
 5582 LeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArg
 CTCTATGGCAATGAGGGCTCGGGGTGGCGGGATGGCTCCTGTCTCCCCGTGGCTCTCGG
 GAGATACCGTTACTCCGACGCCACCCGCCCTACCGAGGACAGAGGGCACCGAGAGCC
 5642 ProSerTrpGlyProThrAspProArgArgSerArgAsnLeuGlyLysValIleAsp
 CCTAGCTGGGGCCCCCACAGACCCCGCGTAGGTGCGCAATTGGTAAGGTCATCGAT
 GGATCGACCCGGGGTGTCTGGGGGCCATCCAGCGCTTAAACCCATTCCAGTAGCTA
 ^ ^
 5650 APAI, 5696 CLAI,
 5702 ThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProLeuValOC AM
 ACCCTTACGTGCGGCTTCGGCACCTCATGGGTACATACCGCTCGTAAATAGTCGAC
 TGGGAATGCACGCCAAGCGGCTGGAGTACCCATGTATGGCGAGCAGATTATCAGCTG
 ^
 5724 HGIE2, 5755 SALI,

FIGURE 22 - Page 1

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn
 2 AGCTTACAAAACAAAATGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAAC
 TCGAATGTTTGTACCGACGTATACGTCGAGTCCCATAATTCCACGATCATGAGTTG
 ^
 1 HIND3, 24 NDEI, 52 SCAI,

 Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp
 62 CCCTCTGTTGCTGCAACACTGGGCTTGGTGCTTACATGTCCAAGGCTCATGGGATCGAT
 GGGAGACAAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCCTAGCTA
 ^
 116 CLAI,

 Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr
 122 CCTAACATCAGGACCGGGGTGAGAACAAATTACCACTGGCACGCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTTGTAAATGGTGACCGTCGGGTAGTGCATGAGGTGG

 Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Ala Tyr Asp Ile Ile Ile Cys
 182 TACGCCAAGTCTCTGCCGACGGCGGGTGCCTCGGGGGCGCTTATGACATAATAATTGT
 ATGCCGTTCAAGGAACGGCTGCCGCCACGAGCCCCCGCGAATACTGTATTAAACA

 Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln
 242 GACGAGTGCCACTCCACGGATGCCACATCCATCTGGCATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTGACAGGAACGGTT

 Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val
 302 GCAGAGACTGGGGGGCGAGACTGGTTGTGCTGCCACCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCGCTCTGACCAACACGAGCGGTGGCGTGGGGAGGCCGAGGCCAG
 ^
 303 ALWN1,

 Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe
 362 ACTGTGCCCATCCAAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTT
 TGACACGGGTAGGGTTGTAGCTCCTCCAAACGAGACAGGTGGTGGCCTCTAGGGAAAA

 Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His
 422 TACGGCAAGGCTATCCCCCTCGAACGTAATCAAGGGGGGAGACATCTCATCTGTGTCAT
 ATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCTCTGTAGAGTAGAACAGAGTA

FIGURE 22 - Page 2

482 SerLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaVal
 TCAAAGAAGAAGTGCACGAACTCGCCGAAAGCTGGTCGATTGGGCATCAATGCCGTG
 AGTTTCTTCTCACGCTGCTGAGCGCGTTCGACCAGCGTAACCGTAGTTACGGCAC

 542 AlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValValVal
 GCCTACTACCGCGGTCTTGACGTGTCGTCATCCCGACCAGCGCGATGTTGTCGTCGTG
 CGGATGATGGCGCCAGAACACTGCACAGGCAGTAGGGCTGGTCGCCCTACAACAGCAC

 550 SAC2, 560 DRD1,

 602 AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsn
 GCAACCGATGCCCTCATGACCGGCTATACCGCGACTTCGACTCCGTGATAGACTGCAAT
 CGTGGCTACGGGAGTACTGGCGATATGCCGCTGAAGCTGAGCCACTATCTGACGTTA

 615 BSPH1,

 662 ThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIle
 ACGTGTGTCACCCAGACAGTCGATTCAGCCTTGACCCCTACCTCACCAATTGAGACAATC
 TGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAG

 722 ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 ACGCTCCCCCAAGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAAG
 TCGCAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCTGACC GTCCCCCTTC

 782 ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 CCAGGCATCTACAGATTGTGCACCGGGGAGCGCCCTCCGGCATGTTGACTCGTCC
 GGTCCGTAGATGTCATAACACCGTGGCCCCCTCGCAGGGGAGGCGTACAAGCTGAGCAGG

 816 BGLI, 833 DRD1,

 842 ValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThr
 GTCCTCTGTGAGTGCATGACGCAGGCTGTGCTTGGTATGAGCTACGCCGAGACT
 CAGGAGACACTACGATACTGCGTCCGACACGAACCATACTCGAGTGCAGGGCGCTCTGA

 881 SAC1,

 902 ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 ACAGTTAGGCTACGAGCGTACATGAACACCCCGGGCTCCCGTGTGCCAGGACCATCTT
 TGTCAATCCGATGCTCGCATGTTGGGGCCCCGAAGGGCACACGGTCTGGTAGAA

 931 SMA1 XMA1,

 962 GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln
 GAATTTGGGAGGGCGTCTTACAGGCCTCACTCATATAGATGCCACTTCTATCCCAG
 CTTAAACCCCTCCCGCAGAAATGTCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTC

 985 STUI,

 1022 ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
 ACAAAAGCAGAGTGGGGAGAACCTTCCTAACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
 TGTTCTGTCACCCCTTTGGAAAGGAATGGACCATCGCATGGTTGGTGGCACACGCGA

 1069 DRA3,

 1082 ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLys
 AGGGCTCAAGCCCCCTCCCCCATCGTGGGACCAGATGTGGAAAGTGTGTTGATTGCCTCAAG

FIGURE 22 - Page 3

TCCCGAGTTGGGGAGGGGGTAGCACCTGGTCTACACCTCACAAACTAACGGAGTT
 ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 1142 CCCACCCCTCCATGGGCCAACACCCCCTGCTATACAGACTGGGCCTGTTCAAGAATGAAATC
 GGGTGGGAGGTACCCGGTGTGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAG
 1150 NCOI,
 ThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluVal
 1202 ACCCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTC
 TGGGACTGCGTGGGTCACTGGTTATGTAGTACTGTACGTACAGCCGGCTGGACCTCCAG
 1230 BSPH1, 1234 DRD1, 1237 AVA3; 1245 EAG1 XMA3, 1250 DRD1,
 ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 1262 GTCACGAGCACCTGGGTGCTCGTGGCGCCTGGCTGCTTGCCGCGTATTGCTG
 CAGTGCTCGTGGACCCACCGAGCAACGCCGCAGGACCGACGAACCGGCCATAACGGAC
 SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIle
 1322 TCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGAAAGCCGGCAATCATA
 AGTGTCCGACGCACCAGTATCACCGTCCCAGCAGAACAGGCCCTCGGCCGTTAGTAT
 1369 NAEI,
 ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
 1382 CCTGACAGGGAAAGTCCTCTACCGAGAGTCGATGAGATGGAAGAGTGCTCTCAGCACTTA
 GGACTGTCCCTTCAGGAGATGGCTCTAACGCTACTCTACCTCTCACGAGAGTCGTGAAT
 1385 DRD1,
 ProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAACGGCCCTCGGCCTC
 GGCATGTAGCTCGTCCCTACTACGAGCGGCTCGTCAAAGTTCGTCTTCCGGAGGCCGGAG
 LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCGTCCCGTCAGGCAGAGGTTATGCCCTGCTGTCCAGACCAACTGGCAA
 GACGTCTGGCGCAGGGCAGTCCGTCTCAAATAGCAGGGACGACAGGTCTGGTTGACCGTT
 1502 PSTI, 1507 TTH3I,
 LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
 1562 AAACTCGAGACCTTCTGGCGAAGCATATGTGGAACCTCATCAGTGGATACAATACTTG
 TTTGAGCTTGGAAAGACCCGCTCGTATAACACCTGAAGTAGTCACCCATGTTATGAAC
 1565 XHOI, 1586 NDEI,
 AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla
 1622 GCGGGCTTGTCAACGCTGCCGGTAACCCGCCATTGCTTCATTGATGGCTTTACAGCT
 CGCCCGAACAGTGGCAGGGACATTGGGGCGGTAAAGTAACACTACCGAAAATGTCGA
 1643 BSTE2, 1677 ALWN1 PVU2,
 AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp
 1682 GCTGTCACCAGCCCCTACCACTAGCCAAACCCCTCCTTCAACATATGGGGGGGTGG
 CGACAGTGGTGGGTGATTGGTATCGGTTGGGAGGAAGTTGTATAACCCCCCACC

FIGURE 22 - Page 4

1742 ValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAla
 GTGGCTGCCAGCTGCCGCCCGGTGCCTACTGCCCTTGTGGCGCTGGCTTAGCT
 CACCGACGGGTCGAGCGGCGGGGCCACGGCGATGACGGAAACACCCGCGACCGAATCGA

1794 ESP1,

1802 GlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyr
 GGCGCCGCATCGGCAGTGTGGACTGGGAAGGTCTCATAGACATCCTTGCAAGGTAT
 CCGCGCGGTAGCCGTACAACCTGACCCCTCCAGGAGTATCTGTAGGAACGTCCCATA

1802 KAS1 NARI,

1862 GlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSer
 GGC CGGGCGTGGCGGGAGCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCC
 CCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTAGTACTCGCCACTCCAGGGAGG

1878 SACI, 1899 BSPH1,

1922 ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly
 ACGGAGGACCTGGTCAATCTACTGCCGCATCCTCTGCCCGAGCCCTCGTAGTCGGC
 TGCCTCTGGACCAGTTAGATGACGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAGCCG

1928 TTH3I,

1982 ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
 GTGGTCTGTGCAGCAATACTGCAGCGCACGTTGGCCCGGGCAGGGGGCAGTGCAGTGG
 CACCAAGACACGTCGTTATGACCGGGCGTGCAACCGGGCCCCGTCACGTCACC

2004 NAEI, 2017 SMAI XMAI,

2042 MetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
 ATGAACCGGCTGATAGCCTCGCCCTCCGGGGAAACCATGTTCCCCCACGCACTACGTG
 TACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGGTGCGTGTGAC

2067 SMAI XMAI, 2093 DRA3,

2102 ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
 CCGGAGAGCGATGCAGCTGCCGCCTACTGCCATACTCAGCAGCCTCACTGTAACCCAG
 GGCCTCTCGCTACGTCGACGGCGCAGTGAACGGTATGAGTCGTCGGAGTGACATTGGGTC

2115 PVU2, 2159 ALWN1,

2162 LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
 CTCCCTGAGGGCAGTCACCAAGTGGATAAGCTGGAGGTGTACCACTCCATGCTCCGGTTCC
 GAGGACTCCCGTGACGTGGTCACCTATTGAGCCTCACATGGTGAGGTACGAGGCCAAGG

2164 MST2, 2220 ECON1,

2222 TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
 TGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTGAGCGACTTTAACGACCTGGCTA
 ACCGATTCCCTGTAGACCCCTGACCTATACTCGCTGAAATTCTGGACCGAT

2282 LysAlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyr
 AAAGCTAACGCTCATGCCACAGCTGCCCTGGATCCCCTTGTGTCCTGCCAGCGCGGGTAT
 TTTCGATTGAGTACGGTGTGACGGACCCTAGGGAAACACAGGACGGTCGCGCCATA

2285 ESP1, 2300 PVU2, 2310 BAMHI,

FIGURE 22 - Page 5

2342 LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 AAGGGGGTCTGGCGAGGGGACGGCATCATGCACACTCGTGCCTACTGTGGAGCTGAGATC
 TTCCCCCAGACCGCTCCCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAG

2402 ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 ACTGGACATGTCAAAACGGGACGATGAGGATCGTGGTCTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCTTGAC

^ ^ ^

2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,

2462 TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla
 TGGAGTGGGACCTCCCCATTAAATGCCCTACACCACGGGCCCTGTACCCCCCTTCCTGCG
 ACCTCACCCCTGGAAGGGTAATTACGGATGTGGTGCCCCGGGACATGGGGGAAGGACGC

^

2480 ASE1, 2497 APAI,

2522 ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln
 CCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACTGTGGAGATAAGGCAG
 GGCTTGATGTGCAAGCGCGATACTCCCACAGACGTCTCCTATGCACCTCTATTCCGTC

^

2553 PSTI,

2582 ValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGln
 GTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTAAATGCCGTGCCAG
 CACCCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTAGATTACGGCACGGT

^

2594 DRA3,

2642 ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro
 GTCCCACGCCCGAATTTCACAGAATTGGACGGGTGCGCCTACATAGGTTGCGCCC
 CAGGGTAGCGGGCTTAAAAAGTGTCTAACCTGCCACCGGGATGTATCCAACGCCGG

2702 ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTGCTGCGGGAGGAGGTATCATTAGACTCCACGAATAACCG
 GGGACGTTGGAACGACGCCCTCCCATAGTAAGTCTCATCCTGAGGTGCTTATGGG

^

2757 HGIE2,

2762 ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
 GTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCGTGTGACGTCCATGCTC
 CATCCCAGCGTTAATGAAACGCTCGGCTTGGCCTGCACGGGACAACCTGCAGGTACGAG

^

2809 AAT2,

2822 ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
 ACTGATCCCTCCCATAAACAGCAGAGGCGGCCGGCGAACGGTTGGCGAGGGGATCACCC
 TGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCAAACCGCTCCCTAGTGGG

^

2850 EAG1 XMA3,

2882 ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
 CCCTCTGTGGCCAGCTCTCGGCTAGCCAGCTATCCGCTCCATCTCAAGGCAACTTGC
 GGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACG

^ ^

2889 BALI, 2903 NHEI,

FIGURE 22 - Page 6

2942 ThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGln
 ACCGCTAACCATGACTCCCCATGGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAG
 TGGCGATTGGTACTGAGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCCGTC
 ^ ^
 2966 ESP1, 2969 SACI,
 GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 3002 GAGATGGGCGGCAACATCACCAAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCC
 CTCTACCCGCCGTGAGTGGTCCAACTCAGTCTTGTGTTCAACCATAAGACCTGAGG
 PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 3062 TTCGATCCGCTTGTGGCGGAGGAGCAGCGGGAGATCTCCGTACCCGAGAAATCCTG
 AAGCTAGGCGAACACCGCCTCCTGCTGCCCTAGAGGCATGGCGTCTTAGGAC
 ^
 3096 BGL2,
 ArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro
 3122 CGGAAGTCTCGGAGATTGCCAGGCCCTGCCGTTGGCGCGCCGGACTATAACCCC
 GCCTTCAGAGCCTCTAACGGGTCGGGACGGGCAAACCCGCGCCGGCTGATATTGGGG
 ^
 3143 ALWN1, 3164 EAG1 XMA3,
 ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAAACCACCTGTGGTCCATGGCTGCCG
 GGCGATCACCTCTGCACCTTTCGGGCTGATGCTGGTGGACACCAGGTACCGACGGG
 ^ ^
 3217 HGIE2, 3229 NCOI,
 LeuProProProLysSerProProValProProProArgLysLysArgThrValValLeu
 3242 CTTCCACCTCCAAAGTCCCCCTCTGTGCCCTCGCCTCGGAAGAAGCGGACGGTGGTCCTC
 GAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGAGCCTTCGCCTGCCACCAGGAG
 ^
 ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer
 3302 ACTGAATCAACCTATCTACTGCCTGGCCAGCTCGCCACCAGAAAGCTTGGCAGCTCC
 TGACTTAGTTGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGG
 ^ ^
 3332 SACI, 3346 HIND3,
 SerThrSerGlyIleThrGlyAspAsnThrThrSerSerGluProAlaProSerGly
 3362 TCAACTCCGGCATTACGGGCACAATACGACAACATCCTCTGAGCCCGCCCTCTGGC
 AGTTGAAGGCCGTAATGCCGCTGTTATGCTGTTAGGAGACTCGGGCGGGAAAGACCG
 ^
 CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGluPro
 3422 TGCCCCCCCAGCTCGACGCTGAGTCCTATTCCCTCATGCCCGCCCTGGAGGGGGAGCCT
 ACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGACCTCCCCCTCGGA
 ^
 3437 EAM11051,
 GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAG
 CCCCTAGGCCTAGAATCGCTGCCAGTACCGAGTGGCCAGTCATCACTCCGGTTGCGCCTC
 ^ ^ ^
 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
 AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla
 3542 GATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGC
 CGGCC
 CTACAGCACACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACCGG

FIGURE 22 - Page 7

3589 DRA3, 3600 SAC2,

3602 AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
GCGGAAGAACAGAAACTGCCATCAATGCACTAAAGCAACTCGTTGCTACGTACCCACAAT
CGCCTTCTTGCTTTGACGGTAGTTACGTGATTGAGCAACGATGCAGTGGTGTAA

3611 ALWN1, 3655 PFLM1,

3662 LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
TTGGTGTATTCCACCACCTCACGCAGTGCCTGCCAAAGGCAGAAGAAAGTCACATTGAC
AACCACATAAGGTGGTGGAGTGCACGAAACGGTTCCGTCTTCAGTGTAAACTG

3681 DRA3,

3722 ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
AGACTGCAAGTTCTGGACAGCATTACCAAGGACGTACTCAAGGAGGTAAAGCAGCGCG
TCTGACGTTCAAGACCTGCGTAATGGCCTGCATGAGTTCTCCAATTGTCGCCGC

3782 SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
TCAAAAGTGAAGGCTAAC TGCTATCCGTAGAGGAAGGCTTGACGCCAGCCCCCACAC
AGTTTCACCCGATTGAACGATAGGCATCTCCTCGAACGTCGGACTGCCGGGTGTG

3816 HIND3,

3842 SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
TCAGCCAAATCCAAGTTGGTTATGGGGCAAAGACGTCCATGCCAGAAAGGCC
AGTCGGTTAGGTTCAAACCAATACCCGTTCTGCAGGCAACGGTACGGTCTTCCGG

3875 AAT2, 3890 BGLI,

3902 ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp
GTAACCCACATCAACTCCGTGTTGGAAAGACCTCTGGAAAGACAATGTAACACCAATAGAC
CATGGGTGTAGTTGAGGCACACCTCTGGAAAGACCTCTGTTACATTGTTATCTG

3962 ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyArgLys
ACTACCACATGGCTAACGAGGTTCTGCGTCCAGCCTGAGAAGGGGGTCGTAAG
TGATGGTAGTACCGATTCTGCTCAAAGACGCAAGTCGGACTCTCCCCCAGCATTC

4022 ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
CCAGCTCGTCTCATCGTGTCCCCGATCTGGCGTGCCTGCGAAAAGATGGCTTG
GGTCGAGCAGAGTAGCACAAGGGCTAGACCCGCACCGCACACGCTTTCTACCGAAC

4082 TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
TACGACGTGGTACAAAGCTCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAATAC
ATGCTGACCAATGTTGAGGGAACCGGACTACCCCTCGAGGATGCCTAAGGTTATG

4142 SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
TCACCAAGGACAGCGGGTTGAATTCTCGTCAAGCGTGGAAAGTCCAAGAAACCCAATG
AGTGGCTGTCGCCAACTTAAGGAGCACGTTGACCTTCAGGTTCTTGGGTACG

4160 ECORI,

4202 GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
GGGTCTCGTATGATAACCCGCTGCTTGACTCCACAGTCAGTGAGAGCGACATCCGTACG
CCCAAGAGCATACTATGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGAGGCATGC

FIGURE 22 - Page 8

4229 DRD1, 4236 ALWN1,

4262 GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
GAGGAGGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCC
CTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGTTGGGGCGCACCGTAGTCAGG

4301 BGLI, 4308 BALI,

4322 LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
CTCACCGAGAGGCTTATGTTGGGGCCCTCTACCAATTCAAGGGGGAGAACTGCGGC
GAGTGGCTCTCCGAAATACAACCCCGGGAGAATGGTTAAGTCCCCCTCTGACGCCG

4345 APA1,

4382 TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
TATCGCAGGTGCGCGCGAGCGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGC
ATAGCGTCCACGGCGCGCTCGCCGATGACTGTTATCGACACCATTGTGGAGTGAACG

4442 TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
TACATCAAGGCCCGGGCAGCCTGTCGAGCCGAGGGCTCCAGGACTGCACCATGCTCGTG
ATGTAGTTCCGGGCCCGTCGGACAGCTCGCGTCCCAGGTCTGACGTGGTACGAGCAC

4452 SMA1 XMA1,

4502 CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
TGTGGCGACGACTTAGCTTATCTGTGAAAGCGCGGGGTCCAGGAGGACGCGGGGAGC
ACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCTCCTGCGCCGCTCG

4508 DRD1, 4511 TTH3I,

4562 LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
CTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCTGGGGACCCCCCACAA
GACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGGGTGTT

4622 ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
CCAGAACATCGACTTGGAGCTATAACATCATGCTCCTCAAACGTGTCAGTCGCCACGAC
GGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTGACAGTCAGCGGGTGCTG

4637 SAC1,

4682 GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
GGCGCTGGAAAAGAGGGCTACTACCTCACCCGTGACCCCTACAACCCCCCTCGCGAGAGCT
CCCGGACCTTCTCCAGATGATGGAGTGGCACTGGGATGTTGGGGAGCGCTCGA

4731 NRUI,

4742 AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
GCGTGGGAGACAGCAAGACACACTCCAGTCATTCTGGCTAGGCAACATAATCATGTTT
CGCACCCCTCTGTCGTTCTGTGAGGTCACTATGACTACTGGTAAAGAAATCGCAGGAATATCGG

4802 AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
GCCCCACACTGTGGCGAGGATGATACTGATGACCCATTCTTAGCGTCTTATAGCC
CGGGGGTGTGACACCCGCTCTACTATGACTACTGGTAAAGAAATCGCAGGAATATCGG

4806 PFLM1, 4807 DRA3,

ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu

FIGURE 22 - Page 9

4862 AGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGAA
TCCCTGGTCGAACCTGTCCGGAGCTAACGCTCTAGATGCCCGGACGATGAGGTATCTT

4893 BGL2,

4922 ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTCACTCCAC
GGTGACCTAGATGGAGGTTAGTAAGTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTG

4954 NCOI,

4982 SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
AGTTACTCTCAGGTGAAATCAAATAGGGTGGCCGCATGCCTCAGAAAATTGGGGTACCG
TCAATGAGAGGTCCACTTAGTTATCCCACCGGCGTACGGAGTCGGTCTTGAACCCATGGC

5015 SPHI, 5035 KPNI,

5042 ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly
CCCTTGCAGCTGGAGACACCGGGCCGGAGCGTCCGGCTAGGCTTCTGGCCAGAGGA
GGGAACGCTCGAACCTCTGTGGCCGGCCTCGCAGGCGCGATCCGAAGACCGGTCTCCT

5064 APAI, 5091 BALI,

5102 GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
GGCAGGGCTGCCATATGTGCAAGTACCTCTTCAACTGGCAGTAAGAACAAAGCTCAA
CCGTCCCACGGTATAACACCGTCATGGAGAAGTTGACCGTCATTCTGTTGAGTT

5113 NDEI,

5162 LeuThrProIleAlaAlaAlaGlyGlnLeuAspLeuSerGlyTrpPheThrAlaGlyTyr
CTCACTCCAATAGCGGCCGCTGGCCAGCTGGACTTGTCCGGCTGGTCACGGCTGGCTAC
GAGTGAGGTTATCGCCGGCGACCGTCGACCTGAACAGGCCGACCAAGTGGCCGACCGATG

5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,

5222 SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
AGCGGGGGAGACATTTACACAGCGTGTCTCATGCCGGCCCCGCTGGATCTGGTTTG
TCGCCCTCTGTAAATAGTGTGACAGAGTACGGCCGGGGGACCTAGACCAAAACG

5240 DRA3,

5282 LeuLeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgMetSerThrAsn
CTACT CCTGCTGCTGCAGGGTAGGCATCTACCTCTCCCCAACGAATGAGCACGAAT
GATGAGGACGAACGACGTCCCCATCCGTAGATGGAGGAGGGGTTGGCTTACTCGTGCTTA

5295 PSTI,

5342 ProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGlnAspValLysPhe
CCTAACCTCAAAGAAAGACCAACGTAACACCAACCGCGCCGAGGACGTCAAGTTC
GGATTGGAGTTCTGGTTGCATTGTGGTTGCCGGCGTCTGCAGTTCAAG

5380 NOTI, 5381 EAG1 XMA3, 5390 AAT2, 5401 SMA1 XMA1.

5402 ProGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeu
CCGGGTGGCGGTCAAGATCGTTGGAGTTACTTGTGCGCGCAGGGGCCCTAGATTG
GGCCCACCGCCAGTCTAGCAACCACCTCAAATGAACAACGGCGCGTCCCCGGGATCTAAC

FIGURE 22 - Page 10

5449 APAI,

GlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnPro
 5462 GGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGTAGACGTCAGCCT
 CCACACGCGCGTGTCTTCTGAAGGCTGCCAGCGTTGGAGCTCCATCTGCAGTCGGA

5467 BSSH2, 5478 XMNI, 5502 XHOI, 5511 AAT2,

IleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpPro
 5522 ATCCCCAAGGCTCGTCGGCCCCAGGGCAGGACCTGGGCTCAGCCGGGTACCCCTGGCCC
 TAGGGGTTCCGAGCAGCCGGCTCCCGTCCGGACCCGAGTCGGGCCATGGGAACCGGG

5548 ALWN1, 5558 ESP1, 5564 SMA1 XMA1, 5568 KPN1,

LeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArg
 5582 CTCTATGGCAATGAGGGCTGGGGTGGCGGGATGGCTCCTGTCTCCCCGTGGCTCTCGG
 GAGATACCGTTACTCCGACGCCAACCGCCCTACCGAGGACAGAGGGCACCGAGAGCC

ProSerTrpGlyProThrAspProArgArgSerArgAsnLeuGlyLysValIleAsp
 5642 CCTAGCTGGGGCCCCACAGACCCCGCGTAGGTCGCGCAATTGGGTAAAGGTATCGAT
 GGATCGACCCCGGGGTGTCTGGGGGCCCATCCAGCGTTAACCCATTCCAGTAGCTA

5650 APAI, 5696 CLAI,

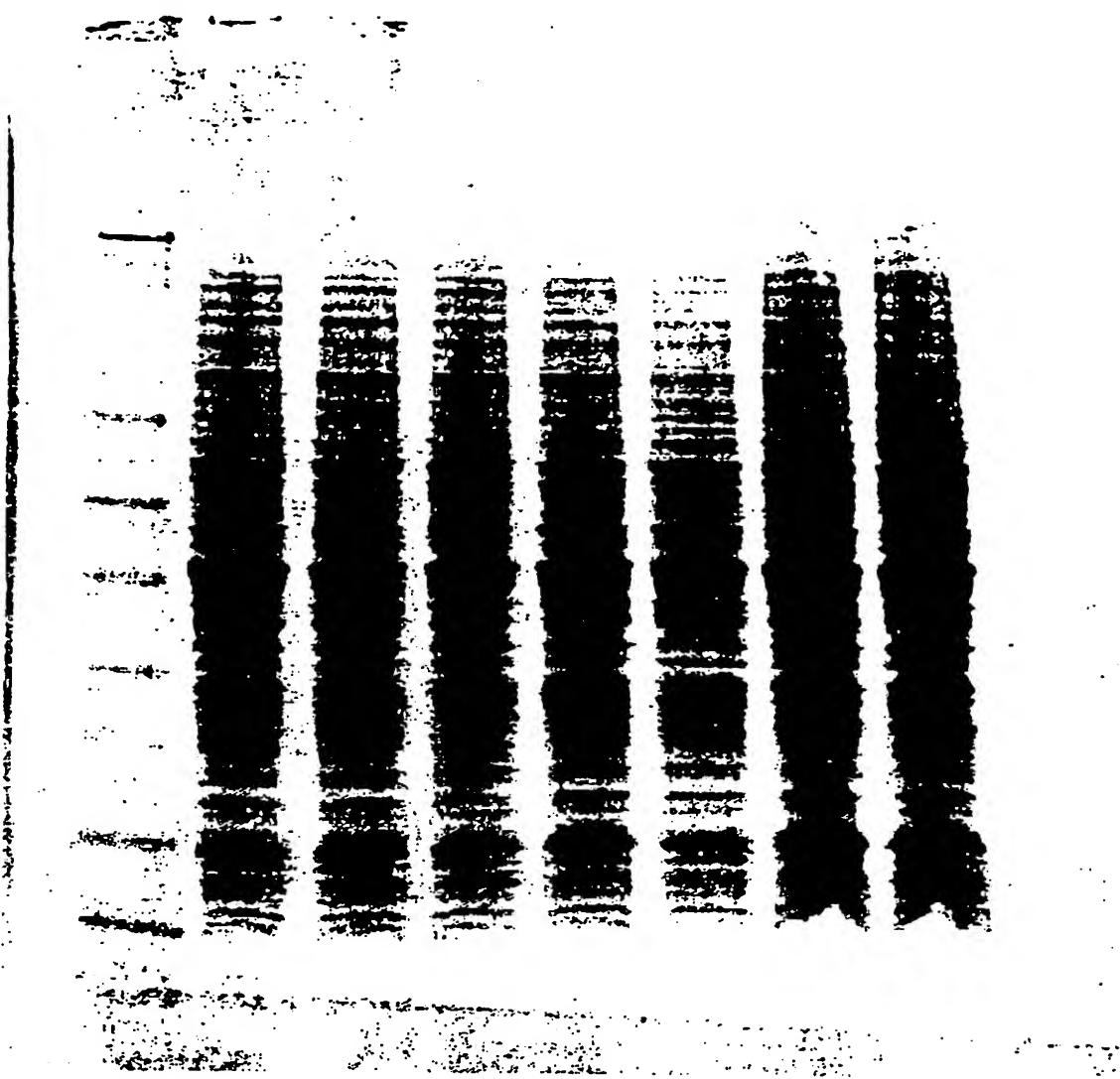
ThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyAlaProLeu
 5702 ACCCTTACGTGCGGCTTCGCCGACCTCATGGGTACATACCGCTCGTCGGGCCCTCTT
 TGGGAATGCACGCCGAAGCGGCTGGAGTACCCCATGTATGGCGAGCAGCCGGGGAGAA

5724 HGIE2, 5750 KAS1 NARI, 5756 ECON1,

GlyGlyAlaAlaArgAlaOC AM
 5762 GGAGGCCGCTGCCAGGGCTTAATAGTCGAC
 CCTCCCGCACGGTCCCGGATTATCAGCTG

5785 SALI,

FIGURE 23



SEQUENCE LISTING

<110> CHIRON CORPORATION et al.

<120> NOVEL HCV NON-STRUCTURAL POLYPEPTIDE

<130> PP01617.003

<140>

<141>

<160> 19

<170> PatentIn Ver. 2.0

<210> 1

<211> 9620

<212> DNA

<213> Artificial Sequence

<220>

<221> CDS

<222> (1990)..(7302)

<220>

<223> Description of Artificial Sequence: Hepatitis C pns345

<400> 1

cgcgcgtttc ggtgatgacg gtgaaaacct ctgacacatg cagctccgg agacggtcac 60
agcttgcgtc taagcggatg ccgggagcag acaagcccgt cagggcgcgt cagcgggtgt 120
tggcggtgt cggggctggc ttaactatgc ggcattcagag cagattgtac tgagagtgc 180
ccatatgaag cttttgcaa aagcctaggc ctccaaaaaa gcctcctcac tacttctgga 240
atagctcaga ggccgaggcg gcctcggct ctgcataaat aaaaaaaaaatt agtcagccat 300
ggggcggaga atggcggaa ctgggggggg agggattat tggctattgg ccattgcata 360
cgttgtatct atatcataat atgtacattt atattggctc atgtccaata tgaccgccat 420
gttgacattt attattgact agttataat agtaatcaat tacggggtca ttagttcata 480
gccccatataat ggagttccgc gttacataac ttacggtaaa tggccgcct ggctgaccgc 540
ccaaacgaccc ccccccattt acgtcaataa tgacgtatgt tcccatagta acgccaatag 600
ggactttcca ttgacgtcaa tgggtggagt attacggta aactgcccac ttggcagtac 660
atcaagtgttca tcataatgcca agtccgcccc ctattgacgt caatgacggt aaatggcccg 720
cctggcatta tgcccaagtac atgaccttac gggactttcc tacttggcag tacatctacg 780
tattagtcat cgcttattacc atgggtatgc ggtttggca gtacaccaat gggcgtggat 840
agcggtttga ctcacgggaa tttccaagtc tccacccat tgacgtcaat gggagttgt 900

tttggcacca aaatcaacgg gactttccaa aatgtcgtaa taaccccggc ccgttgacgc 960
 aaatgggcgg taggcgtgta cggtgggagg tctatataag cagagctcg ttagtgaacc 1020
 gtcagatcgc ctggagacgc catccacgct gtttgacct ccatagaaga caccgggacc 1080
 gatccagcct ccggggccgg gaacggtgca ttgaaacgcg gattccccgt gccaagagtg 1140
 acgtaagtac cgccatataga ctctataggc acacccctt ggctttatg catgtatac 1200
 tgaaaaatgc ttggggccta tacacccccc ctccatgc tataggtgat ggtatagctt 1260
 agcctatagg tggtgggtat tgaccattat tgaccactcc cctattggtg acgataactt 1320
 ccattactaa tccataacat ggctcttgc cacaactatc tctattggct atatgccaat 1380
 actctgtcct tcagagactg acacggactc tgtatTTTA caggatgggg tccatttt 1440
 atttacaaat tcacatatac aacaacggcg tccccgtgc ccgcagttt tattaaacat 1500
 agcgtggat ctccgacatc tcgggtacgt gttccggaca tgggctttc tccggtagcg 1560
 gcggagcttc cacatccgag ccctggtccc atccgtccag cggctcatgg tcgctcggca 1620
 gtccttgct cctaacagtg gaggccagac ttaggcacag cacaatgccc accaccacca 1680
 gtgtgccgca caaggccgtg gcggtagggt atgtgtctga aaatgagctc ggagattggg 1740
 ctcgcacctg gacgcagatg gaagacttaa ggcagcggca gaagaagatg caggcagctg 1800
 agttgtttaa ttctgataag agtcagaggt aactcccggt gcggtagctgt taacggtgaa 1860
 gggcagtgtt gtcgtggcag tactcggtgc tgccgcgcgc gccaccagac ataatacg 1920
 acagactaac agactgttcc tttccatggg tctttctgc agtcaccgtc gtcgaccaa 1980
 gaattcacc atg gct gca tat gca gct cag ggc tat aag gtg cta gta ctc 2031
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu
 1 5 10

aac	ccc	tct	gtt	gct	gca	aca	ctg	ggc	ttt	ggt	gct	tac	atg	tcc	aag	2079
Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	Gly	Phe	Gly	Ala	Tyr	Met	Ser	Lys	
15	20														30	

gct	cat	ggg	atc	gat	cct	aac	atc	agg	acc	ggg	gtg	aga	aca	att	acc	2127
Ala	His	Gly	Ile	Asp	Pro	Asn	Ile	Arg	Thr	Gly	Val	Arg	Thr	Ile	Thr	
35	40														45	

act	ggc	agc	ccc	atc	acg	tac	tcc	acc	tac	ggc	aag	ttc	ctt	gcc	gac	2175
Thr	Gly	Ser	Pro	Ile	Thr	Tyr	Ser	Thr	Tyr	Gly	Lys	Phe	Leu	Ala	Asp	
50	55														60	

ggc	ggg	tgc	tgc	ggg	ggc	gct	tat	gac	ata	ata	att	tgt	gac	gag	tgc	2223
Gly	Gly	Cys	Ser	Gly	Gly	Ala	Tyr	Asp	Ile	Ile	Ile	Cys	Asp	Glu	Cys	
65	70														75	

cac	tcc	acg	gat	gcc	aca	tcc	atc	ttg	ggc	att	ggc	act	gtc	ctt	gac	2271
His	Ser	Thr	Asp	Ala	Thr	Ser	Ile	Leu	Gly	Ile	Gly	Thr	Val	Leu	Asp	
80	85														90	

caa gca gag act gcg ggg ggc aga ctg gtt gtg ctc gcc acc gcc acc	2319
Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr	
95 100 105 110	
cct ccg ggc tcc gtc act gtg ccc cat ccc aac atc gag gag gtt gct	2367
Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala	
115 120 125	
ctg tcc acc acc gga gag atc cct ttt tac ggc aag gct atc ccc ctc	2415
Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu	
130 135 140	
gaa gta atc aag ggg ggg aga cat ctc atc ttc tgt cat tca aag aag	2463
Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys	
145 150 155	
aag tgc gac gaa ctc gcc gca aag ctg gtc gca ttg ggc atc aat gcc	2511
Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala	
160 165 170	
gtg gcc tac tac cgc ggt ctt gac gtg tcc gtc atc ccc acc agc ggc	2559
Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly	
175 180 185 190	
gat gtt gtc gtc gtg gca acc gat gcc ctc atg acc ggc tat acc ggc	2607
Asp Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly	
195 200 205	
gac ttc gac tcg gtg ata gac tgc aat acg tgt gtc acc cag aca gtc	2655
Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val	
210 215 220	
gat ttc agc ctt gac cct acc ttc acc att gag aca atc acg ctc ccc	2703
Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro	
225 230 235	
caa gat gct gtc tcc cgc act caa cgt cgg ggc agg act ggc agg ggg	2751
Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly	
240 245 250	
aag cca ggc atc tac aga ttt gtg gca ccc ggg gag cgc ccc tcc ggc	2799
Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly	
255 260 265 270	
atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gca ggc tgt gct	2847
Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala	
275 280 285	
tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta cga gcg tac	2895
Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr	
290 295 300	
atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat ctt gaa ttt tgg	2943
Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp	
305 310 315	
gag ggc gtc ttt aca ggc ctc act cat ata gat gcc cac ttt cta tcc	2991
Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser	
320 325 330	

cag aca aag cag agt ggg gag aac ctt cct tac ctg gta gcg tac caa Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln 335 340 345 350	3039
gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg tgg gac cag Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln 355 360 365	3087
atg tgg aag tgt ttg att cgc ctc aag ccc acc ctc cat ggg cca aca Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr 370 375 380	3135
ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa atc acc ctg acg Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr 385 390 395	3183
cac cca gtc acc aaa tac atc atg aca tgc atg tcg gcc gac ctg gag His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu 400 405 410	3231
gtc gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg gct gct ttg Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu 415 420 425 430	3279
gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata gtg ggc agg gtc Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val 435 440 445	3327
gtc ttg tcc ggg aag ccg gca atc ata cct gac agg gaa gtc ctc tac Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr 450 455 460	3375
cga gag ttc gat gag atg gaa gag tgc tct cag cac tta ccg tac atc Arg Glu Phe Asp Glu Met Glu Cys Ser Gln His Leu Pro Tyr Ile 465 470 475	3423
gag caa ggg atg atg ctc gcc gag cag ttc aag cag aag gcc ctc ggc Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly 480 485 490	3471
ctc ctg cag acc gcg tcc cgt cag gca gag gtt atc gcc cct gct gtc Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val 495 500 505 510	3519
cag acc aac tgg caa aaa ctc gag acc ttc tgg gcg aag cat atg tgg Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp 515 520 525	3567
aac ttc atc agt ggg ata caa tac ttg gcg ggc ttg tca acg ctg cct Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro 530 535 540	3615
ggg aac ccc gcc att gct tca ttg atg gct ttt aca gct gct gtc acc Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr 545 550 555	3663
agc cca cta acc act agc caa acc ctc ctc ttc aac ata ttg ggg ggg Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly 560 565 570	3711

tgg gtg gct gcc cag ctc gcc gcc ccc ggt gcc gct act gcc ttt gtg Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val 575 580 585 590	3759
ggc gct ggc tta gct ggc gcc atc ggc agt gtt gga ctg ggg aag Gly Ala Gly Leu Ala Gly Ala Ile Gly Ser Val Gly Leu Gly Lys 595 600 605	3807
gtc ctc ata gac atc ctt gca ggg tat ggc gcg ggc gtg gcg gga gct Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala 610 615 620	3855
ctt gtg gca ttc aag atc atg agc ggt gag gtc ccc tcc acg gag gac Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp 625 630 635	3903
ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc gga gcc ctc gta gtc Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val 640 645 650	3951
ggc gtg gtc tgt gca gca ata ctg cgc cg ^g cac gtt ggc ccg ggc gag Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu 655 660 665 670	3999
ggg gca gtg cag tgg atg aac cgg ctg ata gcc ttc gcc tcc cg ^g ggg Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly 675 680 685	4047
aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat gca gct gcc Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala 690 695 700	4095
cgc gtc act gcc ata ctc agc agc ctc act gta acc cag ctc ctg agg Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg 705 710 715	4143
cga ctg cac cag tgg ata agc tcg gag tgt acc act cca tgc tcc ggt Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly 720 725 730	4191
tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg ttg agc gac Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp 735 740 745 750	4239
ttt aag acc tgg cta aaa gct aag ctc atg cca cag ctg cct ggg atc Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile 755 760 765	4287
ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg cga ggg gac Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp 770 775 780	4335
ggc atc atg cac act cgc tgc cac tgt gga gct gag atc act gga cat Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His 785 790 795	4383
gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc tgc agg aac Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn 800 805 810	4431

atg tgg agt ggg acc ttc ccc att aat gcc tac acc acg ggc ccc tgt Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys 815 820 825 830	4479
acc ccc ctt cct gcg ccg aac tac acg ttc gcg cta tgg agg gtg tct Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser 835 840 845	4527
gca gag gaa tac gtg gag ata agg cag gtg ggg gac ttc cac tac gtg Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val 850 855 860	4575
acg ggt atg act act gac aat ctt aaa tgc ccg tgc cag gtc cca tcg Thr Gly Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser 865 870 875	4623
ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cat agg ttt gcg Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala 880 885 890	4671
ccc ccc tgc aag ccc ttg ctg cgg gag gag gta tca ttc aga gta gga Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly 895 900 905 910	4719
ctc cac gaa tac ccg gta ggg tcgcaa tta cct tgc gag ccc gaa ccg Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro 915 920 925	4767
gac gtg gcc gtg ttg acg tcc atg ctc act gat ccc tcc cat ata aca Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr 930 935 940	4815
gca gag gcg gcc ggg cga agg ttg gcg agg gga tca ccc ccc tct gtg Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val 945 950 955	4863
gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc aag gca act Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr 960 965 970	4911
tgc acc gct aac cat gac tcc cct gat gct gag ctc ata gag gcc aac Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn 975 980 985 990	4959
ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg gtt gag tca Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser 995 1000 1005	5007
gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt gtg gcg gag Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu 1010 1015 1020	5055
gag gac gag cgg gag atc tcc gta ccc gca gaa atc ctg cgg aag tct Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser 1025 1030 1035	5103
cggtt gac ggg gac tcc gtt tgg ggc cgg ccg gac tat aac Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn 1040 1045 1050	5151

ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa cca cct gtg	5199
Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val	
1055 1060 1065 1070	
gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct gtg cct ccg	5247
Val His Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro Val Pro Pro	
1075 1080 1085	
cct cgg aag aag cgg acg gtg gtc ctc act gaa tca acc cta tct act	5295
Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr	
1090 1095 1100	
gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc tca act tcc	5343
Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser	
1105 1110 1115	
ggc att acg ggc gac aat acg aca aca tcc tct gag ccc gcc cct tct	5391
Gly Ile Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser	
1120 1125 1130	
ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc atg ccc ccc	5439
Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro	
1135 1140 1145 1150	
ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg tca tgg tca	5487
Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser	
1155 1160 1165	
acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc tgc tca atg	5535
Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met	
1170 1175 1180	
tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc gcg gaa gaa	5583
Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu	
1185 1190 1195	
cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta cgt cac cac	5631
Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His	
1200 1205 1210	
aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa agg cag aag	5679
Asn Leu Val Tyr Ser Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys	
1215 1220 1225 1230	
aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat tac cag gac	5727
Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp	
1235 1240 1245	
gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag gct aac ttg	5775
Val Leu Lys Glu Val Lys Ala Ala Ser Lys Val Lys Ala Asn Leu	
1250 1255 1260	
cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac tca gcc aaa	5823
Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys	
1265 1270 1275	
tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat gcc aga aag	5871
Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys	
1280 1285 1290	

gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg gaa gac aat Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn 1295 1300 1305 1310	5919
gta aca cca ata gac act acc atc atg gct aag aac gag gtt ttc tgc Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys 1315 1320 1325	5967
gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc atc gtg ttc Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe 1330 1335 1340	6015
ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg tac gac gtg Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val 1345 1350 1355	6063
gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac gga ttc caa Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln 1360 1365 1370	6111
tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg tgg aag tcc Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser 1375 1380 1385 1390	6159
aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc ttt gac tcc Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser 1395 1400 1405	6207
aca gtc act gag agc gac atc cgt acg gag gag gca atc tac caa tgt Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys 1410 1415 1420	6255
tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc ctc acc gag Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu 1425 1430 1435	6303
agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg gag aac tgc Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys 1440 1445 1450	6351
ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act agc tgt ggt Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly 1455 1460 1465 1470	6399
aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt cga gcc gca Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala 1475 1480 1485	6447
ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac tta gtc gtt Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val 1490 1495 1500	6495
atc tgt gaa agc gcg ggg gtc cag gag gac gcg agc ctg aga gcc Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala 1505 1510 1515	6543
ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg gac ccc cca Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro 1520 1525 1530	6591

caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc tcc aac gtg 6639
 Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val
 1535 1540 1545 1550

 tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac ctc acc cgt 6687
 Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg
 1555 1560 1565

 gac cct aca acc ccc ctc gcg aga gct gcg tgg gag aca gca aga cac 6735
 Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His
 1570 1575 1580

 act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt gcc ccc aca 6783
 Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr
 1585 1590 1595

 ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc gtc ctt ata 6831
 Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile
 1600 1605 1610

 gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc tac ggg gcc 6879
 Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala
 1615 1620 1625 1630

 tgc tac tcc ata gaa cca ctg gat cta cct cca atc att caa aga ctc 6927
 Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu
 1635 1640 1645

 cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca ggt gaa atc 6975
 His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile
 1650 1655 1660

 aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg ccc ttg cga 7023
 Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg
 1665 1670 1675

 gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt ctg gcc aga 7071
 Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg
 1680 1685 1690

 gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac tgg gca gta 7119
 Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val
 1695 1700 1705 1710

 aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc cag ctg gac 7167
 Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp
 1715 1720 1725

 ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac att tat cac 7215
 Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His
 1730 1735 1740

 agc gtg tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc cta ctc ctg 7263
 Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu
 1745 1750 1755

 ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga tgaaggttgg 7312
 Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1760 1765 1770

ggtaaacact ccggcctaaa aaaaaaaaaa aatctagaaa ggcgcgccaa gatatcaagg 7372
atccactacg cgtagagct cgctgatcg cctcgactgt gccttctagt tgccagccat 7432
ctgttgtttg cccctccccc gtgccttcct tgaccctgga aggtgccact cccactgtcc 7492
tttcctaata aaatgaggaa attgcatcgc attgtctgag taggtgtcat tctattctgg 7552
gggggtgggt gggcaggac agcaaggggg aggattggga agacaatagc aggcattgtg 7612
gggagctctt ccgcttcctc gctcaactgac tcgctgcgct cggcggttcg gctgcggcga 7672
gcggtatcag ctcaactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca 7732
ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaaa ggccgcgttg 7792
ctggcggttt tccataggct ccggccccc gacgagcatc acaaaaatcg acgctcaagt 7852
cagaggtggc gaaacccgac aggactataa agataccagg cggttcccccc tggaagctcc 7912
ctcgtgcgct ctccgttcc gaccctgccc cttaccggat acctgtccgc ctttctccct 7972
tcgggaagcg tggcgcttcc tcaatgctca cgctgttaggt atctcagttc ggtgttaggtc 8032
gttcgctcca agctggctg tgtgcaccaa ccccccgttc agccccgaccg ctgcgcctta 8092
tccggtaact atcgtcttga gtccaaacccg gtaagacacg acttatacgcc actggcagca 8152
gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag 8212
tggtggccta actacggcta cactagaagg acagtatttgc gatctgcgc tctgctgaag 8272
ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaac caccgctgg 8332
agcggtggtt ttttgggtt caagcagcag attacgcgca gaaaaaaaaagg atctcaagaa 8392
gatccttga tctttctac ggggtctgac gctcagtggaa acgaaaactc acgttaaggg 8452
attttgtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga 8512
agttttaat caatctaaag tatatatgag taaacttggt ctgacagttt ccaatgctta 8572
atcagtgagg cacctatctc agcgatctgt ctatccgtt catccatagt tgcctgactc 8632
cccgctgtgt agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg 8692
ataccgcgag acccacgctc accggctcca gatttacatcg caataaacca gccagccgga 8752
agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt 8812
tgccgggaag cttagagtaag tagttcgcca gttaatagtt tgcgcaacgt tggtgccatt 8872
gctacaggca tcgtgggtgc acgctcgatcg tttggatgg cttcattcag ctccgggttcc 8932
caacgatcaa ggcgagttac atgatcccc atgttgcgca aaaaagcggt tagctccttc 8992
ggtcctccga tcgttgcag aagtaagttg gccgcagtgt tatcactcat gttatggca 9052
gcactgcata attctttaac tgtcatgcca tccgtaagat gttttctgt gactggtgag 9112

tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgccccg 9172
tcaatacggg ataataccgc gccacatagc agaactttaa aagtgtcat cattggaaaa 9232
cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa 9292
cccactcgtg caccacaactg atcttcagca tctttactt tcaccagcgt ttctgggtga 9352
gcaaaaacag gaaggcaaaa tgccgcaaaa aagggataa gggcgacacg gaaatgttga 9412
atactcatac tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg 9472
agcggataca tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt 9532
ccccgaaaag tgcccacctga cgtctaagaa accattatta tcatgacatt aacctataaa 9592
aataqqcqta tcacqqaqccttgcqtc 9620

<210> 2
<211> 1771
<212> PRT
<213> Hepatitis C virus

<220>
<223> Description of Artificial Sequence: Hepatitis C pns345

<400> 2
Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys Cys His Ser
 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys
145 150 155 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala

165	170	175
Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val		
180	185	190
Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe		
195	200	205
Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe		
210	215	220
Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp		
225	230	235
240		
Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro		
245	250	255
Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe		
260	265	270
Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr		
275	280	285
Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn		
290	295	300
Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly		
305	310	315
320		
Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr		
325	330	335
Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr		
340	345	350
Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp		
355	360	365
Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu		
370	375	380
Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro		
385	390	395
400		
Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val		
405	410	415
Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala		
420	425	430
Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu		
435	440	445
Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu		
450	455	460
Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln		
465	470	475
480		

Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu
 485 490 495

 Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
 500 505 510

 Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe
 515 520 525

 Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
 530 535 540

 Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
 545 550 555 560

 Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575

 Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590

 Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605

 Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
 610 615 620

 Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
 625 630 635 640

 Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655

 Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
 660 665 670

 Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
 675 680 685

 Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
 690 695 700

 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
 705 710 715 720

 His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735

 Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750

 Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765

 Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
 770 775 780

 Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
 785 790 795 800

Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
 805 810 815
 Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
 820 825 830
 Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
 835 840 845
 Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
 850 855 860
 Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
 865 870 875 880
 Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
 885 890 895
 Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910
 Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
 915 920 925
 Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
 930 935 940
 Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
 945 950 955 960
 Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975
 Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
 980 985 990
 Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
 995 1000 1005
 Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
 1010 1015 1020
 Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
 1025 1030 1035 1040
 Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
 1045 1050 1055
 Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070
 Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
 1075 1080 1085
 Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
 1090 1095 1100
 Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
 105 1110 1115 1120

Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135
 Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
 1140 1145 1150
 Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
 1155 1160 1165
 Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
 1170 1175 1180
 Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 1185 1190 1195 1200
 Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
 1205 1210 1215
 Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
 1220 1225 1230
 Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
 1235 1240 1245
 Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
 1250 1255 1260
 Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 1265 1270 1275 1280
 Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
 1285 1290 1295
 Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310
 Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325
 Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 1330 1335 1340
 Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
 1345 1350 1355 1360
 Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
 1365 1370 1375
 Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
 1380 1385 1390
 Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
 1395 1400 1405
 Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
 1410 1415 1420
 Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
 1425 1430 1435 1440

Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
 1445 1450 1455
 Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470
 Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
 1475 1480 1485
 Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
 1490 1495 1500
 Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
 505 1510 1515 1520
 Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535
 Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
 1540 1545 1550
 Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
 1555 1560 1565
 Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
 1570 1575 1580
 Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
 585 1590 1595 1600
 Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
 1605 1610 1615
 Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
 1620 1625 1630
 Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
 1635 1640 1645
 Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
 1650 1655 1660
 Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
 665 1670 1675 1680
 Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695
 Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
 1700 1705 1710
 Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp Leu Ser
 1715 1720 1725
 Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
 1730 1735 1740
 Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala
 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
1765 1770

<210> 3
<211> 9620

<212> DNA

<213> Artificial Sequence

<220>

<221> CDS

<222> (1990)..(7302)

<220>

<223> Description of Artificial Sequence: pDeltaNS3NS5

<400> 3

cgcgcgttcc ggtgatgacg gtgaaaacct ctgacacatg cagctccgg agacggtcac 60
agtttgtctg taagcggatg ccgggagcag acaagcccgt cagggcgcgt cagcgggtgt 120
tggcggtgt cggggctggc ttaactatgc ggcatcagag cagattgtac tgagagtgc 180
ccatatgaag cttttgcaa aagcctaggc ctccaaaaaaaaa gcctcctcac tacttctgga 240
atagctcaga ggccgaggcg gcctcgccct ctgcataaat aaaaaaaaaatt agtcagccat 300
ggggcggaga atggcggaa ctgggcgggg agggattat tggctattgg ccattgcata 360
cggtgtatct atatcataat atgtacattt atattggctc atgtccaata tgaccgccat 420
gttgacattt attattgact agttattaat agtaatcaat tacggggtca ttagttcata 480
gcccatatat ggagttccgc gttacataac ttacggtaaa tggccgcct ggctgaccgc 540
ccaacgaccc ccccccattt acgtcaataa tgacgtatgt tcccatagta acgccaatag 600
ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac ttggcagttac 660
atcaagtgtt tcatatgcca agtccgcccc ctattgacgt caatgacgtt aaatggcccg 720
cctggcatta tgcccaattt acgtacattt gggactttcc tacttggcag tacatctacg 780
tattagtcat cgctattacc atggtgatgc ggtttggca gtacaccaat gggcgtggat 840
agcgggttga ctcacgggaa tttcaagtc tccacccat tgacgtcaat gggagttgt 900
tttggcacca aaatcaacgg gactttccaa aatgtcgtaa taacccgc ccttgcacgc 960
aaatggcgg taggcgtgtt cgggtggagg tctatataag cagagctgtt ttagtgaacc 1020
gtcagatcgc ctggagacgc catccacgtt gttttgaccc tccatagaaga caccggacc 1080
gatccagccct ccgcggccgg gaacggtgca ttggAACGCG gattccccgt gccaagagt 1140
acgtaagtac cgcctataga ctctataggc acaccccttt ggctttagtgc catgtataac 1200
tggggggccta tacaccccg ctccttatgc tatagggtat ggtatagctt 1260

agcctatagg tgtgggttat tgaccattat tgaccactcc cctattggtg acgatacttt 1320
 ccattactaa tccataacat ggctcttgc cacaactatc tctattggct atatgccaat 1380
 actctgtcct tcagagactg acacggactc tgtatTTTA caggatgggg tccatttt 1440
 atttacaaat tcacatatac aacaacgccc tccccgtgc ccgcagttt tattaaacat 1500
 agcgtggat ctccgacatc tcgggtacgt gttccggaca tgggtcttc tccggtagcg 1560
 gcggagcttc cacatccgag ccctggtccc atccgtccag cggctcatgg tcgctcggca 1620
 gctccttgct cctaacagtg gaggccagac ttaggcacag cacaatgccc accaccacca 1680
 gtgtgccgca caaggccgtg gcggtagggt atgtgtctga aaatgagctc ggagattggg 1740
 ctcgcacctg gacgcagatg gaagacttaa ggcagcggca gaagaagatg caggcagctg 1800
 agttgttgta ttctgataag agtcagaggt aactcccggt gcegtgctgt taacgggtgga 1860
 gggcagtgtta gtctgagcag tactcggtgc tgccgcgcgc gccaccagac ataatacgctg 1920
 acagactaac agactgttcc tttccatggg tctttctgc agtcaccgtc gtcgaccaa 1980
 gaattcacc atg gct gca tat gca gct cag ggc tat aag gtg cta gta ctc 2031
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu
 1 5 10

aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac atg tcc aag 2079
 Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys
 15 20 25 30

gct cat ggg atc gat cct aac atc agg acc ggg gtg aga aca att acc 2127
 Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr
 35 40 45

act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc ctt gcc gac 2175
 Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp
 50 55 60

ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt gac gag tgc 2223
 Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys
 65 70 75

cac tcc acg gat gcc aca tcc atc ttg ggc att ggc act gtc ctt gac 2271
 His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp
 80 85 90

caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc gcc acc gcc acc 2319
 Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr
 95 100 105 110

cct ccg ggc tcc gtc act gtg ccc cat ccc aac atc gag gag gtt gct 2367
 Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala
 115 120 125

ctg tcc acc acc gga gag atc cct ttt tac ggc aag gct atc ccc ctc 2415
 Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu
 130 135 140

gaa gta atc aag ggg qgg aga cat ctc atc ttc tgt cat tca aag aag Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys 145 150 155	2463
aag tgc gac gaa ctc gcc gca aag ctg gtc gca ttg ggc atc aat gcc Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala 160 165 170	2511
gtg gcc tac tac cgc ggt ctt gac gtg tcc gtc atc ccg acc agc ggc Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly 175 180 185 190	2559
gat gtt gtc gtc gtg gca acc gat gcc ctc atg acc ggc tat acc ggc Asp Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly 195 200 205	2607
gac ttc gac tcg gtg ata gac tgc aat acg tgt gtc acc cag aca gtc Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val 210 215 220	2655
gat ttc agc ctt gac cct acc ttc acc att gag aca atc acg ctc ccc Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro 225 230 235	2703
caa gat gct gtc tcc cgc act caa cgt cgg ggc agg act ggc agg ggg Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly 240 245 250	2751
aag cca ggc atc tac aga ttt gtg gca ccg ggg gag cgc ccc tcc ggc Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly 255 260 265 270	2799
atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gca ggc tgt gct Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala 275 280 285	2847
tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta cga gcg tac Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr 290 295 300	2895
atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat ctt gaa ttt tgg Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp 305 310 315	2943
gag ggc gtc ttt aca ggc ctc act cat ata gat gcc cac ttt cta tcc Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser 320 325 330	2991
cag aca aag cag agt ggg gag aac ctt cct tac ctg gta gcg tac caa Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln 335 340 345 350	3039
gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg tgg gac cag Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln 355 360 365	3087
atg tgg aag tgt ttg att cgc ctc aag ccc acc ctc cat ggg cca aca Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr 370 375 380	3135

ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa atc acc ctg acg 3183
 Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr
 385 390 395

 cac cca gtc acc aaa tac atc atg aca tgc atg tcg gcc gac ctg gag 3231
 His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu
 400 405 410

 gtc gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg gct gct ttg 3279
 Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu
 415 420 425 430

 gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata gtg ggc agg gtc 3327
 Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val
 435 440 445

 gtc ttg tcc ggg aag ccg gca atc ata cct gac agg gaa gtc ctc tac 3375
 Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr
 450 455 460

 cga gag ttc gat gag atg gaa gag tgc tct cag cac tta ccg tac atc 3423
 Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile
 465 470 475

 gag caa ggg atg atg ctc gag cag ttc aag cag aag gcc ctc ggc 3471
 Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly
 480 485 490

 ctc ctg cag acc gcg tcc cgt cag gca gag gtt atc gcc cct gct gtc 3519
 Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val
 495 500 505 510

 cag acc aac tgg caa aaa ctc gag acc ttc tgg gcg aag cat atg tgg 3567
 Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp
 515 520 525

 aac ttc atc agt ggg ata caa tac ttg gcg ggc ttg tca acg ctg cct 3615
 Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro
 530 535 540

 ggt aac ccc gcc att gct tca ttg atg gct ttt aca gct gct gtc acc 3663
 Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr
 545 550 555

 agc cca cta acc act agc caa acc ctc ctc ttc aac ata ttg ggg ggg 3711
 Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly
 560 565 570

 tgg gtg gct gcc cag ctc gcc ccc ggt gcc gct act gcc ttt gtg 3759
 Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val
 575 580 585 590

 ggc gct ggc tta gct ggc gcc atc ggc agt gtt gga ctg ggg aag 3807
 Gly Ala Gly Leu Ala Gly Ala Ile Gly Ser Val Gly Leu Gly Lys
 595 600 605

 gtc ctc ata gac atc ctt gca ggg tat ggc gcg ggc gtg gcg gga gct 3855
 Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala
 610 615 620

ctt gtg gca ttc aag atc atg agc ggt gag gtc ccc tcc acg gag gac Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp 625 630 635	3903
ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc gga gcc ctc gta gtc Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val 640 645 650	3951
ggc gtg gtc tgt gca gca ata ctg cgc cg ^g cac gtt ggc ccg ggc gag Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu 655 660 665 670	3999
ggg gca gtg cag tgg atg aac cgg ctg ata gcc ttc gcc tcc cgg ggg Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly 675 680 685	4047
aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat gca gct gcc Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala 690 695 700	4095
cgc gtc act gcc ata ctc agc agc ctc act gta acc cag ctc ctg agg Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg 705 710 715	4143
cga ctg cac cag tgg ata agc tcg gag tgt acc act cca tgc tcc ggt Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly 720 725 730	4191
tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg ttg agc gac Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp 735 740 745 750	4239
ttt aag acc tgg cta aaa gct aag ctc atg cca cag ctg cct ggg atc Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile 755 760 765	4287
ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg cga ggg gac Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp 770 775 780	4335
ggc atc atg cac act cgc tgc cac tgt gga gct gag atc act gga cat Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His 785 790 795	4383
gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc tgc agg aac Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn 800 805 810	4431
atg tgg agt ggg acc ttc ccc att aat gcc tac acc acg ggc ccc tgt Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys 815 820 825 830	4479
acc ccc ctt cct gcg ccg aac tac acg ttc gcg cta tgg agg gtg tct Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser 835 840 845	4527
gca gag gaa tac gtg gag ata agg cag gtg ggg gac ttc cac tac gtg Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val 850 855 860	4575

acg ggt atg act act gac aat ctt aaa tgc ccg tgc cag gtc cca tcg 4623
 Thr Gly Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser
 865 870 875

 ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cat agg ttt gcg 4671
 Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala
 880 885 890

 ccc ccc tgc aag ccc ttg ctg cgg gag gag gta tca ttc aga gta gga 4719
 Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly
 895 900 905 910

 ctc cac gaa tac ccg gta ggg tcg caa tta cct tgc gag ccc gaa ccg 4767
 Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro
 915 920 925

 gac gtg gcc gtg ttg acg tcc atg ctc act gat ccc tcc cat ata aca 4815
 Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr
 930 935 940

 gca gag gcg gcc ggg cga agg ttg gcg agg gga tca ccc ccc tct gtg 4863
 Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val
 945 950 955

 gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc aag gca act 4911
 Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr
 960 965 970

 tgc acc gct aac cat gac tcc cct gat gct gag ctc ata gag gcc aac 4959
 Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn
 975 980 985 990

 ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg gtt gag tca 5007
 Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser
 995 1000 1005

 gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt gtg gcg gag 5055
 Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu
 1010 1015 1020

 gag gac gag cgg gag atc tcc gta ccc gca gaa atc ctg cgg aag tct 5103
 Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser
 1025 1030 1035

 cgg aga ttc gcc cag gcc ctg ccc gtt tgg gcg cgg ccg gac tat aac 5151
 Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn
 1040 1045 1050

 ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa cca cct gtg 5199
 Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val
 1055 1060 1065 1070

 gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct gtg cct ccg 5247
 Val His Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro Val Pro Pro
 1075 1080 1085

 cct cgg aag aag cgg acg gtg gtc ctc act gaa tca acc cta tct act 5295
 Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr
 1090 1095 1100

gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc tca act tcc Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser 1105 1110 1115	5343 .
ggc att acg ggc gac aat acg aca aca tcc tct gag ccc gcc cct tct Gly Ile Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser 1120 1125 1130	5391
ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc atg ccc ccc Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro 1135 1140 1145 1150	5439
ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg tca tgg tca Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser 1155 1160 1165	5487
acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc tgc tca atg Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met 1170 1175 1180	5535
tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc gcg gaa gaa Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu 1185 1190 1195	5583
cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta cgt cac cac Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His 1200 1205 1210	5631
aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa agg cag aag Asn Leu Val Tyr Ser Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys 1215 1220 1225 1230	5679
aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat tac cag gac Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp 1235 1240 1245	5727
gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag gct aac ttg Val Leu Lys Glu Val Lys Ala Ala Ser Lys Val Lys Ala Asn Leu 1250 1255 1260	5775
cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac tca gcc aaa Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys 1265 1270 1275	5823
tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat gcc aga aag Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys 1280 1285 1290	5871
gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg gaa gac aat Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn 1295 1300 1305 1310	5919
gta aca cca ata gac act acc atc atg gct aag aac gag gtt ttc tgc Val Thr Pro Ile Asp Thr Ile Met Ala Lys Asn Glu Val Phe Cys 1315 1320 1325	5967
gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc atc gtg ttc Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe 1330 1335 1340	6015

ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg tac gac gtc 6063
 Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val
 1345 1350 1355

gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac gga ttc caa 6111
 Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln
 1360 1365 1370

tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg tgg aag tcc 6159
 Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser
 1375 1380 1385 1390

aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc ttt gac tcc 6207
 Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser
 1395 1400 1405

aca gtc act gag agc gac atc cgt acg gag gag gca atc tac caa tgt 6255
 Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys
 1410 1415 1420

tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc ctc acc gag 6303
 Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu
 1425 1430 1435

agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg gag aac tgc 6351
 Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys
 1440 1445 1450

ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act agc tgt ggt 6399
 Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly
 1455 1460 1465 1470

aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt cga gcc gca 6447
 Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala
 1475 1480 1485

ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac tta gtc gtt 6495
 Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val
 1490 1495 1500

atc tgt gaa agc gcg ggg gtc cag gag gac gcg gcg agc ctg aga gcc 6543
 Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala
 1505 1510 1515

ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg gac ccc cca 6591
 Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro
 1520 1525 1530

caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc tcc aac gtg 6639
 Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val
 1535 1540 1545 1550

tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac ctc acc cgt 6687
 Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg
 1555 1560 1565

gac cct aca acc ccc ctc gcg aga gct gcg tgg gag aca gca aga cac 6735
 Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His
 1570 1575 1580

act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt gcc ccc aca	6783
Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr	
1585 1590 1595	
ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc gtc ctt ata	6831
Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile	
1600 1605 1610	
gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc tac ggg gcc	6879
Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala	
1615 1620 1625 1630	
tgc tac tcc ata gaa cca ctg gat cta cct cca atc att caa aga ctc	6927
Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu	
1635 1640 1645	
cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca ggt gaa atc	6975
His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile	
1650 1655 1660	
aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg ccc ttg cga	7023
Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg	
1665 1670 1675	
gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt ctg gcc aga	7071
Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg	
1680 1685 1690	
gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac tgg gca gta	7119
Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val	
1695 1700 1705 1710	
aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc cag ctg gac	7167
Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp	
1715 1720 1725	
ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac att tat cac	7215
Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His	
1730 1735 1740	
agc gtg tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc cta ctc ctg	7263
Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu	
1745 1750 1755	
ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga tgaaggttgg	7312
Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg	
1760 1765 1770	
ggtaaacact ccggcctaaa aaaaaaaaaa aatctagaaa ggccgcgcca gatatcaagg	7372
atccactacg cgtagatcg cgctgatcg cctcgactgt gccttctagt tgccagccat	7432
ctgttgtttg cccctccccc gtgccttcct tgaccctgga aggtgccact cccactgtcc	7492
tttcctaata aaatgaggaa attgcattgc attgtctgag tagtgtcat tctattctgg	7552
gggttggtt gggcaggac agcaaggggg aggattggga agacaatagc aggcattgtg	7612
gggagcttt ccgttccctc gctcaactgac tcgctgcgtcg cggtcggtcg gctgcggcga	7672

gcggtatcag ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca 7732
ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaaa ggccgcgtt 7792
ctggcgcccccc tccataggct ccgcggccct gacgagcatc acaaaaatcg acgctcaagt 7852
cagaggtggc gaaaccgcac aggactataa agataccagg cgtttcccccc tggaagctcc 7912
tcgtgcgtct ctccgttcc gaccctgccg cttaccggat acctgtccgc ctttctccct 7972
tcggaaagcg tggcgcttcc tcaatgctca cgctgttaggt atctcagttc ggtgttaggtc 8032
gttcgctcca agctggctg tgtgcaccaa ccccccgttc agccccgaccc ctgcgcctta 8092
tccggtaact atcgtcttga gtccaaacccg gtaagacacg acttatcgcc actggcagca 8152
gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag 8212
tggtggccta actacggcta cactagaagg acagtatttgc gatatctgcgc tctgctgaag 8272
ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaaacaac caccgctgg 8332
agcgggtggtt tttttgttttcaagcagcag attacgcgca gaaaaaaaaagg atctcaagaa 8392
gatcctttaa tcttttctac ggggtctgac gctcagtggaa acgaaaactc acgttaaggg 8452
attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga 8512
agttttaaat caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta 8572
atcagtgagg cacctatctc agcgatctgt ctatccgtt catccatagt tgcctgactc 8632
cccgctgtgt agataactac gatacggag ggcttaccat ctggccccag tgctgcaatg 8692
ataccgcgag acccacgctc accggctcca gatttatcag caataaacca gccagccgga 8752
agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt 8812
tgccgggaag cttagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt 8872
gctacaggca tcgtgggtgc acgctcgctg tttggatgg cttcattcag ctccgggttcc 8932
caacgatcaa ggcgagttac atgatcccc atgttgtca aaaaagcggt tagctccttc 8992
ggtcctccga tcgttgcag aagtaagttg gcccgcgtgt tatcactcat gttatggca 9052
gcactgcata attcttttac tgtcatgcca tccgtaagat gctttctgt gactggtag 9112
tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgcgc ttgcccggcg 9172
tcaatacggg ataataccgc gccacatagc agaactttaa aagtgcgtcat cattggaaaa 9232
cgttcttcgg ggcgaaaaact ctcaaggatc ttaccgtgt tgagatccag ttgcgtatgtaa 9292
cccactcgatc caccctactg atcttcagca tcctttactt tcaccagcgt ttctgggtga 9352
gcaaaaacag gaaggcaaaa tgccgcaaaa aagggataa gggcgacacg gaaatgttga 9412
atactcatac tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg 9472

agcggataca tatttgaatg tatTTAGAAA aataaACAAA taggggttcc gcgacattt 9532
 ccccGAAAAG tgccacctga cgtctaAGAA accATTATTA tcATGACATT AACCTATAAA 9592
 aataggcgtA tcacgaggcc ctTTCGTC 9620

<210> 4
 <211> 1771
 <212> PRT
 <213> Artificial Sequence

<400> 4
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
 1 5 10 15
 Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
 20 25 30
 Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
 35 40 45
 Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
 50 55 60
 Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser
 65 70 75 80
 Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
 85 90 95
 Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
 100 105 110
 Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
 115 120 125
 Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
 130 135 140
 Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys
 145 150 155 160
 Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala
 165 170 175
 Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
 180 185 190
 Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe
 195 200 205
 Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe
 210 215 220
 Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp
 225 230 235 240
 Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro

245	250	255
Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe		
260	265	270
Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr		
275	280	285
Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn		
290	295	300
Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly		
305	310	315
Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr		
325	330	335
Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr		
340	345	350
Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp		
355	360	365
Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu		
370	375	380
Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro		
385	390	395
400		
Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val		
405	410	415
Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala		
420	425	430
Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu		
435	440	445
Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu		
450	455	460
Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln		
465	470	475
480		
Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu		
485	490	495
Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr		
500	505	510
Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe		
515	520	525
Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn		
530	535	540
Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro		
545	550	555
560		

Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575
 Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590
 Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605
 Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
 610 615 620
 Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
 625 630 635 640
 Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655
 Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
 660 665 670
 Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
 675 680 685
 Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Arg Val
 690 695 700
 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
 705 710 715 720
 His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735
 Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750
 Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765
 Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
 770 775 780
 Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
 785 790 795 800
 Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
 805 810 815
 Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
 820 825 830
 Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
 835 840 845
 Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
 850 855 860
 Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
 865 870 875 880

Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
 885 890 895

 Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910

 Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
 915 920 925

 Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
 930 935 940

 Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
 945 950 955 960

 Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975

 Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
 980 985 990

 Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
 995 1000 1005

 Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
 1010 1015 1020

 Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
 1025 1030 1035 1040

 Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
 1045 1050 1055

 Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070

 Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
 1075 1080 1085

 Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
 1090 1095 1100

 Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
 1095 1110 1115 1120

 Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135

 Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
 1140 1145 1150

 Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
 1155 1160 1165

 Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
 1170 1175 1180

 Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 1185 1190 1195 1200

Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
 1205 1210 1215
 Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
 1220 1225 1230
 Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
 1235 1240 1245
 Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
 1250 1255 1260
 Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 1265 1270 1275 1280
 Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
 1285 1290 1295
 Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310
 Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325
 Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 1330 1335 1340
 Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
 1345 1350 1355 1360
 Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
 1365 1370 1375
 Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
 1380 1385 1390
 Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
 1395 1400 1405
 Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
 1410 1415 1420
 Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
 1425 1430 1435 1440
 Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
 1445 1450 1455
 Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470
 Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
 1475 1480 1485
 Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
 1490 1495 1500
 Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
 1505 1510 1515 1520

Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535
 Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
 1540 1545 1550
 Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
 1555 1560 1565
 Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
 1570 1575 1580
 Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
 585 1590 1595 1600
 Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
 1605 1610 1615
 Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
 1620 1625 1630
 Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
 1635 1640 1645
 Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
 1650 1655 1660
 Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
 665 1670 1675 1680
 Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695
 Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
 1700 1705 1710
 Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp Leu Ser
 1715 1720 1725
 Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
 1730 1735 1740
 Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala
 745 1750 1755 1760
 Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1765 1770

<210> 5
 <211> 4282
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: pCMVII

<400> 5
 tcgcgcgttt cggatgac ggtaaaaacc tctgacacat gcagctcccg gagacggtca 60

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaaactatg cgccatcaga gcagattgtt ctgagagtgc 180
accatatgaa gcttttgca aaaggcttagg cctccaaaaa agcctcctca ctacttctgg 240
aatagctcag agggcgaggc ggcctcggcc tctgcataaa taaaaaaaaat tagtcagcca 300
tggggcggag aatgggcgga actgggcggg gagggattt ttggctattt gccattgcat 360
acgttgtatc tatatcataa tatgtacatt tatattggct catgtccaaat atgaccgcca 420
tggttgcatt gattatttgc tagttattaa tagtaatcaa ttacggggtc attagttcat 480
agccccatata tggagttccg cgttacataa cttacggtaa atggcccgc tggctgaccg 540
cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata 600
gggactttcc attgacgtca atgggtggag tatttacggt aaactgccc cttggcagta 660
catcaagtgt atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc 720
gcctggcatt atgcccagta catgaccta cgggactttc ctacttggca gtacatctac 780
gtattagtca tcgctattac catggtgatg cggttttggc agtacaccaa tggcgtgga 840
tagcggtttt actcacgggg atttccaagt ctccacccca ttgacgtcaa tggagttt 900
ttttggcacc aaaatcaacg ggactttcca aaatgtcgta ataacccgc cccgttgacg 960
caaatggcg gtaggcgtgt acgggtggag gtctatataa gcagagctcg tttagtgaac 1020
cgtcagatcg cctggagacg ccattccacgc tggtttgacc tccatagaag acaccggac 1080
cgatccagcc tccgcggccg ggaacgggtgc attggaacgc ggattcccg tgccaagagt 1140
gacgtaaatgta cccctatag actctatagg cacacccctt tggcttttat gcatgctata 1200
ctgttttgg cttggggcct atacacccca gtttcttat gctataggtt atggatagc 1260
ttagcctata ggtgtgggtt attgaccatt attgaccact cccctatttt tgacgatact 1320
ttccattact aatccataac atggctctt gccacaacta tctctattttt ctatatgcca 1380
atactctgtc cttcagagac tgacacggac tctgtatTT tacaggatgg ggtccattt 1440
attatttaca aattcacata tacaacaacg ccgtcccccgg tgcccgcaat ttttattaaa 1500
catagcgtgg gatctccacg cgaatctcggt gtacgtgttc cggacatggg ctcttctccg 1560
gtagcggcgg agcttccaca tccgagccct ggtcccatgc ctccagggc tcatggtcgc 1620
tcggcagctc cttgctccta acagtggagg ccagacttag gcacagcaca atgcccacca 1680
ccaccagggtt gcccacaag gccgtggcgg tagggatgt gtctgaaaat gagctcggag 1740
attgggctcg caccgctgac gcagatggaa gacttaaggc agcggcagaa gaagatgcag 1800
gcagctgagt tgggttattt tgataagagt cagaggttaac tcccggttgcg gtgctgtttaa 1860

cgggtggaggg cagtgttagtc tgagcagtagc tcgttgctgc cgcgccgc accagacata 1920
atagctgaca gactaacaga ctgttccccc ccatgggtct tttctgcagt caccgtcgac 1980
gacctaagaa ttcaagactcg agcaagtcta gaaaggcgcg ccaagatatac aaggatccac 2040
tacgcgttag agctcgctga tcagcctcga ctgtgccttc tagttgccag ccatctgttg 2100
tttgccttc cccccgtgcct tccttgaccc tggaagggtgc cactcccact gtccttcct 2160
aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt ctggggggtg 2220
gggtggggca ggacagcaag ggggaggatt gggaaagacaa tagcaggcat gctggggagc 2280
tcttcgcctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg gcgaggcgta 2340
tcagctcact caaaggcggt aatacggtt tccacagaat caggggataa cgccaggaaag 2400
aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg 2460
ttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcaaggagg 2520
tggcgaaacc cgacaggact ataaagatac cagggcttcc cccctggaag ctccctcgta 2580
cgctctcctg ttccgaccct gccgcttacc ggataacctgt ccgcctttct cccttcggga 2640
agcgtggcgc tttctcaatg ctcacgctgt aggtatctca gttcggtgt a gtcgttcgc 2700
tccaagctgg gctgtgtgca cgaacccccc gttcagcccg accgctgcgc cttatccgg 2760
aactatcgcc ttgagtccaa cccggtaaga cacgacttat cgccactggc agcagccact 2820
ggtaaacagga ttagcagagc gaggtatgtt ggcggtgcta cagagttctt gaagtgggtgg 2880
cctaactacg gctacactag aaggacagta tttggtatct gcgcctgtc gaagccagtt 2940
accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggt 3000
ggtttttttgg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct 3060
ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtt aaggattttg 3120
gtcatgagat tatcaaaaag gatcttacc tagatcctt taaattaaaa atgaagtttt 3180
aaatcaatct aaagtatata ttagttaact tggctgtaca gttaccaatg cttaatcagt 3240
gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg actccccgtc 3300
gtgttagataa ctacgatacg ggagggctta ccatctggcc ccagtgcgtc aatgataccg 3360
cgagacccac gctcacccggc tccagattta tcagcaataa accagccagc cggaaggggcc 3420
gagcgcagaa gtggctctgc aactttatcc gcctccatcc agtctattaa ttgttgcggg 3480
gaagcttagag taagtagttc gccagttaat agtttgcgc acgttggc cattgtaca 3540
ggcatcgtaggg tgcacgctc gtcgtttggatggcttcat tcagctccgg ttcccaacga 3600
tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggttagctc ctccggcct 3660

ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcattgttat ggcagcactg 3720
 cataattctc ttactgtcat gccatccgta agatgtttt ctgtgactgg tgagtactca 3780
 accaagtcat tctgagaata gtgtatgccc cgaccgagtt gtccttgcgg ggcgtcaata 3840
 cgggataata ccgcgccaca tagcagaact ttaaaaagtgc tcattcattgg aaaacgttct 3900
 tcggggcgaa aactctcaag gatcttacccg ctgtttagat ccagttcgat gtaaccact 3960
 cgtgcaccca actgatcttc agcatcttt actttcacca gcgtttctgg gtgagcaaaa 4020
 acaggaaggc aaaatgccgc aaaaaaggaa ataagggcga cacggaaatg ttgaatactc 4080
 atactcttcc ttttcaata ttattgaagc atttacagg gttattgtct catgagcgga 4140
 tacatatttg aatgtattta gaaaaataaaa caaatagggg ttccgcgcac atttccccga 4200
 aaagtgccac ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg 4260
 cgtatcacga ggccctttcg tc 4282

<210> 6
 <211> 6299
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: pNS34a

<220>
 <221> CDS
 <222> (1990)..(4047)

<400> 6
 cgcgcgttcc ggtgatgacg gtgaaaacct ctgacacatg cagctccgg agacggcac 60
 agcttgtctg taagcggatg ccgggagcag acaagccgt cagggcgcgt cagcgggtgt 120
 tggcgggtgt cggggctggc ttaactatgc ggcacatcagag cagattgtac tgagagtgc 180
 ccatatgaag cttttgcaa aagcctaggc ctccaaaaaaaa gcctcctcac tacttctgga 240
 atagctcaga ggccgaggcg gcctcggcct ctgcataaat aaaaaaaaaatt agtcagccat 300
 ggggcggaga atggcggaa ctgggcgggg agggaaattat tggctattgg ccattgcata 360
 cgttgtatct atatcataat atgtacattt atattggctc atgtccaata tgaccgccat 420
 gttgacattt attattgact agttataat agtaatcaat tacggggtca ttagttcata 480
 gcccataatggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc 540
 ccaacgaccc ccgcatttgc acgtcaataa tgacgtatgt tcccatagta acgccaatag 600
 ggactttcca ttgacgtcaa tgggtggagt attacggta aactgcccac ttggcagttac 660
 atcaagtgtatcataatgcata agtccgcggcccttattgacgtt caatgacggtaaatggcccg 720

cctggcatta tgcccagtac atgaccttac gggactttcc tacttggcag tacatctacg 780
 tattagtcat cgctattacc atggtgatgc gggtttggca gtacaccaat gggcggtggat 840
 agcggttga ctcacgggaa tttccaagtc tccaccccat tgacgtcaat gggagtttgt 900
 tttggcacca aaatcaacgg gactttccaa aatgtcgtaa taaccccgcc ccgttgacgc 960
 aaatgggcgg taggcgtgta cgggtggagg tctatataag cagagctcg ttagtgaacc 1020
 gtcagatcgc ctggagacgc catccacgct gtttgacct ccatagaaga caccgggacc 1080
 gatccagcct ccgcggccgg gaacggtgca ttggaacgcg gattccccgt gccaagagtg 1140
 acgtaagtac cgccatataga ctctataggc acacccctt ggctcttatg catgctatac 1200
 tgttttggc ttggggccta tacacccctg ctccttatgc tataagggtat ggtatagttt 1260
 agcctatagg tgtgggttat tgaccattat tgaccactcc cctattggtg acgatacttt 1320
 ccattactaa tccataacat ggctcttgc cacaactatc tctattggct atatgcaat 1380
 actctgtcct tcagagactg acacggactc tgtatTTTA caggatgggg tccatttt 1440
 atttacaaat tcacatatac aacaacgccc tccccgtgc ccgcagttt tattaaacat 1500
 agcgtgggat ctccgacatc tcgggtacgt gttccggaca tgggctcttc tccggtagcg 1560
 gcggagcttc cacatccgag ccctggtccc atccgtccag cggctcatgg tcgctcggca 1620
 gtccttgct cctaacagtg gaggccagac ttaggcacag cacaatgccc accaccacca 1680
 gtgtgccgca caaggccgtg gcggtagggt atgtgtctga aaatgagctc ggagattggg 1740
 ctcgcacctg gacgcagatg gaagacttaa ggcagcggca gaagaagatg caggcagctg 1800
 agttgtttaa ttctgataag agtcagaggt aactcccgaa gtcggtagctgt taacgggttga 1860
 gggcagtgtttaa gtctgagcag tactcggttc tgccgcgcg gccaccagac ataatacg 1920
 acagactaac agactgttcc tttccatggg tctttctgc agtcacccgtc gtcgaccaa 1980
 gaattcacc atg gcg ccc atc acg gcg tac gcc cag cag aca agg ggc ctc 2031
 Met Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu
 1 5 10

cta	ggg	tgc	ata	atc	acc	agc	cta	act	ggc	cg	gac	aaa	aac	caa	gtg	2079
Leu	Gly	Cys	Ile	Ile	Thr	Ser	Leu	Thr	Gly	Arg	Asp	Lys	Asn	Gln	Val	
15				20				25						30		

gag	gg	gt	g	tc	c	ag	tt	gt	tca	act	g	cc	ca	ac	tt	ct	g	ca	2127
Glu	Gly	Glu	Val	Gln	Ile	Val	Ser	Thr	Ala	Ala	Gln	Thr	Phe	Leu	Ala				
															35	40	45		

acg	tgc	atc	aat	ggg	gt	tg	tgc	tgg	act	gtc	tac	cac	ggg	gcc	gga	acg	2175	
Thr	Cys	Ile	Asn	Gly	Val	Cys	Trp	Thr	Val	Tyr	His	Gly	Ala	Gly	Thr			
															50	55	60	

agg	acc	atc	g	c	g	t	c	cc	a	ag	g	gt	c	c	at	g	at	acc	aat	2223
-----	-----	-----	---	---	---	---	---	----	---	----	---	----	---	---	----	---	----	-----	-----	------

Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn	65	70	75	
gta gac caa gac ctt gtg ggc tgg ccc gct tcg caa ggt acc cgc tca				2271
Val Asp Gln Asp Leu Val Gly Trp Pro Ala Ser Gln Gly Thr Arg Ser	80	85	90	
ttg aca ccc tgc act tgc ggc tcc tcg gac ctt tac ctg gtc acg agg				2319
Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg	95	100	105	
cac gcc gat gtc att ccc gtg cgc cgg cgg ggt gat agc agg ggc agc				2367
His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser	115	120	125	
ctg ctg tcg ccc cgg ccc att tcc tac ttg aaa ggc tcc tcg ggg ggt				2415
Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly	130	135	140	
ccg ctg ttg tgc ccc gcg ggg cac gcc gtg ggc ata ttt agg gcc gcg				2463
Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Ile Phe Arg Ala Ala	145	150	155	
gtg tgc acc cgt gga gtg gct aag gcg gtg gac ttt atc cct gtg gag				2511
Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu	160	165	170	
aac cta gag aca acc atg agg tcc ccg gtg ttc acg gat aac tcc tct				2559
Asn Leu Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser	175	180	185	
cca cca gta gtg ccc cag agc ttc cag gtg gct cac ctc cat gct ccc				2607
Pro Pro Val Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro	195	200	205	
aca ggc agc ggc aaa agc acc aag gtc ccg gct gca tat gca gct cag				2655
Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln	210	215	220	
ggc tat aag gtg cta gta ctc aac ccc tct gtt gct gca aca ctg ggc				2703
Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly	225	230	235	
ttt ggt gct tac atg tcc aag gct cat ggg atc gat cct aac atc agg				2751
Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg	240	245	250	
acc ggg gtg aga aca att acc act ggc agc ccc atc acg tac tcc acc				2799
Thr Gly Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr	255	260	265	
tac ggc aag ttc ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac				2847
Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp	275	280	285	
ata ata att tgt gac gag tgc cac tcc acg gat gcc aca tcc atc ttg				2895
Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu	290	295	300	

ggc att ggc act gtc ctt gac caa gca gag act gcg ggg gcg aga ctg Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu 305 310 315	2943
gtt gtg ctc gcc acc gcc acc cct ccg ggc tcc gtc act gtg ccc cat Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His 320 325 330	2991
ccc aac atc gag gag gtt gct ctg tcc acc acc gga gag atc cct ttt Pro Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe 335 340 345 350	3039
tac ggc aag gct atc ccc ctc gaa gta atc aag ggg ggg aga cat ctc Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu 355 360 365	3087
atc ttc tgt cat tca aag aag aag tgc gac gaa ctc gcc gca aag ctg Ile Phe Cys His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu 370 375 380	3135
gtc gca ttg ggc atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg Val Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val 385 390 395	3183
tcc gtc atc ccg acc agc ggc gat gtt gtc gtc gtg gca acc gat gcc Ser Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala 400 405 410	3231
ctc atg acc ggc tat acc ggc gac ttc gac tcg gtg ata gac tgc aat Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn 415 420 425 430	3279
acg tgt gtc acc cag aca gtc gat ttc acg ctt gac cct acc ttc acc Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr 435 440 445	3327
att gag aca atc acg ctc ccc caa gat gct gtc tcc cgc act caa cgt Ile Glu Thr Ile Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg 450 455 460	3375
cgg ggc agg act ggc agg ggg aag cca ggc atc tac aga ttt gtg gca Arg Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala 465 470 475	3423
ccg ggg gag cgc ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu 480 485 490	3471
tgc tat gac gca ggc tgt gct tgg tat gag ctc acg ccc gcc gag act Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr 495 500 505 510	3519
aca gtt agg cta cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc Thr Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys 515 520 525	3567
cag gac cat ctt gaa ttt tgg gag ggc gtc ttt aca ggc ctc act cat Gln Asp His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His 530 535 540	3615

ata gat gcc cac ttt cta tcc cag aca aag cag agt ggg gag aac ctt	3663
Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu	
545 550 555	
cct tac ctg gta gcg tac caa gcc acc gtg tgc gct agg gct caa gcc	3711
Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala	
560 565 570	
cct ccc cca tcg tgg gac cag atg tgg aag tgt ttg att cgc ctc aag	3759
Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys	
575 580 585 590	
ccc acc ctc cat ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt	3807
Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val	
595 600 605	
cag aat gaa atc acc ctg acg cac cca gtc acc aaa tac atc atg aca	3855
Gln Asn Glu Ile Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr	
610 615 620	
tgc atg tcg gcc gac ctg gag gtc gtc acg agc acc tgg gtg ctc gtt	3903
Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val	
625 630 635	
ggc ggc gtc ctg gct ttg gcc gcg tat tgc ctg tca aca ggc tgc	3951
Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys	
640 645 650	
gtg gtc ata gtg ggc agg gtc ttg tcc ggg aag ccg gca atc ata	3999
Val Val Ile Val Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile	
655 660 665 670	
cct gac agg gaa gtc ctc tac cga gag ttc gat gag atg gaa gag tgc	4047
Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys	
675 680 685	
 taggatccac tacgcgttag agctcgctga tcagcctcga ctgtgccttc tagttgccag 4107	
 ccatctgttg tttgccccctc cccccgtgcct tccttgaccc tggaagggtgc cactcccact 4167	
 gtcctttcct aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt 4227	
 ctgggggtg ggggtggggca ggacagcaag ggggaggatt gggaaagacaa tagcaggcat 4287	
 gctggggagc tcttcgcctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg 4347	
 gcgagcggta tcagctcaact caaaggcggt aatacggta tccacagaat cagggataa 4407	
 cgcagggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc 4467	
 gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc 4527	
 aagtcaaggagg tggcgaaacc cgacaggact ataaagatac caggcggttc cccctggaag 4587	
 ctccctcggt cgctctcctg ttccgaccct gccgcttacc ggataccctgt ccgcctttct 4647	
 cccttcggga agcgtggcgc tttctcaatg ctcacgctgt aggtatctca gttcggtgta 4707	
 ggtcgttcgc tccaaagctgg gctgtgtgca cgaacccccc gttcagcccg acccgctgcgc 4767	

cttatccggtaactatcgctttgagtcacccggtaaga caccgacttat cgccactggc 4827
agcagccact ggtaacagga ttagcagagc gaggtatgtt ggcggtgcta cagagttctt 4887
gaagtggtgg ccttaactacg gctacactag aaggacagta tttggtatct gcgcctctgct 4947
gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc 5007
tggtagcggt ggttttttt tttgcaagca gcagattacg cgcaaaaaaaa aaggatctca 5067
agaagatcct ttgatctttt ctacggggtc tgacgctcag tggAACGAAA actcacgtta 5127
agggattttt gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 5187
atgaagtttt aaatcaatct aaagtatata tgagtaaact tggcttgaca gttaccaatg 5247
cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg 5307
actccccgtc gtgtagataa ctacgataacg ggagggctta ccatctggcc ccagtgcgc 5367
aatgataccg cgagacccac gtcaccggc tccagattt tcagcaataa accagccagc 5427
cggaaggggcc gagcgcagaa gtggcctgc aactttatcc gcctccatcc agtctattaa 5487
ttgttgcgg gaagctagag taagtagttc gccagttat agttgcgc acgttgttgc 5547
cattgctaca ggcacatcggtt tgacgctc gtcgtttggat atggcttcat tcagctccgg 5607
ttcccaacga tcaaggcgag ttacatgatc cccatgttgc tgcaaaaaag cggttagctc 5667
cttcggcctt ccgatcggtt tcagaagtaa gttggccgca gtgttatcac tcattgttat 5727
ggcagcactg cataattctc ttactgtcat gccatccgtt agatcctttt ctgtgactgg 5787
tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctctgccc 5847
ggcgtcaata cgggataata ccgcgcacaca tagcagaact ttaaaaagtgc tcattgttgg 5907
aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 5967
gttaaccact cgtgcaccca actgatcttc agcatctttt accttcacca gcgttctgg 6027
gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggaa ataagggcga cacggaaatg 6087
ttgaataactc atactcttcc ttttcaata ttattgttgc atttatcagg gttattgtct 6147
catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac 6207
atttccccga aaagtgcacac ctgacgtctt agaaaccatt attatcatga cattaaccta 6267
taaaaatagg cgtatcacga ggccctttcg tc 6299

<210> 7
<211> 686
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: pNS34a

<400> 7

Met Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly
 1 5 10 15

Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
 20 25 30

Glu Val Gln Ile Val Ser Thr Ala Ala Gln Thr Phe Leu Ala Thr Cys
 35 40 45

Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr
 50 55 60

Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp
 65 70 75 80

Gln Asp Leu Val Gly Trp Pro Ala Ser Gln Gly Thr Arg Ser Leu Thr
 85 90 95

Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
 100 105 110

Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
 115 120 125

Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
 130 135 140

Leu Cys Pro Ala Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys
 145 150 155 160

Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu
 165 170 175

Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro
 180 185 190

Val Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly
 195 200 205

Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr
 210 215 220

Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly
 225 230 235 240

Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly
 245 250 255

Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly
 260 265 270

Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile
 275 280 285

Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile
 290 295 300

Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val
 305 310 315 320

Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn
 325 330 335

Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly
 340 345 350

Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe
 355 360 365

Cys His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala
 370 375 380

Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val
 385 390 395 400

Ile Pro Thr Ser Gly Asp Val Val Val Ala Thr Asp Ala Leu Met
 405 410 415

Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys
 420 425 430

Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu
 435 440 445

Thr Ile Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly
 450 455 460

Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly
 465 470 475 480

Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr
 485 490 495

Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val
 500 505 510

Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp
 515 520 525

His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp
 530 535 540

Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr
 545 550 555 560

Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro
 565 570 575

Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr
 580 585 590

Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn
 595 600 605

Glu Ile Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met
 610 615 620

Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly
 625 630 635 640

Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val
 645 650 655

Ile Val Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp
 660 665 670

Arg Glu Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys
 675 680 685

<210> 8
<211> 19912
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: pd.deltaNS3NS5

<220>
<221> CDS
<222> (12745)..(18057)

<400> 8
atcgatcccta ccccttgcgc taaaagaagta tatgtgccta ctaacgcttg tctttgtctc 60
tgtcaactaaa cactggattt ttactcccag atacttattt tggactaattt taaatgattt 120
cgatatcaacg ttcttaatat cgctgaatct tccacaattt atgaaaagtag ctatggagag 180
gaattggtat aaagtttttg tttttgtaaa tctcgaagta tactcaaacg aattttagat 240
tttctcagtg atctcccaga tgctttcacc ctcacttaga agtgctttaa gcatttttt 300
actgtggcta ttcccttat ctgcttccttc cgatgattcg aactgttaattt gcaaactact 360
tacaatataca gtgatatacg attgatgttt ttgtccatag taaggataaa ttgtaaattc 420
ccaagcagga atcaatttctt ttaatgaggc ttccagaattt gttgttttt gcgtcttgta 480
tttaaactgg agtgattttat tgacaatatac gaaactcagc gaattgctta tgatagtatt 540
atagctcatg aatgtggctc tcttgattgc tggccatgc tggtaatca tccaaacataa 600
ataggtagt tcagcagcac ataatgctat tttctcacctt gaaggctttt caaaccttcc 660
cacaaactga cgaacaagca ccttaggtgg tggatccat aatataatcaa attgtggcat 720
gcttagcgcc gatcttgcgtt gcaattgata tctagttca actactctat ttatcttgta 780
tcttgcagta ttcaaacacg ctaactcgaa aaactaactt taattgtcctt gttgtctcg 840
cggtatcttc atttcatattt ttaaaaatgc acctttgctg ctttcctta attttagac 900
ggccccgcagg ttctgtttgc ggtactatct tggataaaaa agttgttttg acatgtgatc 960
ggccccgcagg ttctgtttgc ggtactatct tggataaaaa agttgttttg acatgtgatc 1020

tgcacagatt ttataatgta ataagcaaga atacattatc aaacgaacaa tactggtaaa 1080
agaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140
cttatacgct ctggatgta tgcggctgt ttattgaaat gattgctcct gatgtagata 1200
ttgatataga gagtaaacgt aagtctgatg agctactctt tccaggatat gtcataaggc 1260
ccatggaaatc tctcacaacc ggtaggccgt atggcttga ttctagcgca gaagattcca 1320
gcgtatcttc tgactccagt gctgaggtaa tttgcctgc tgcgaagatg gttaaggaaa 1380
ggtttgattc gattggaaat ggtatgctct cttcacaaga agcaagtcag gctgccatag 1440
atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500
ctataataat aggaagattg cccgagaaaag acaagaagag agctaccgaa atgctcatga 1560
aaaaaatgga ttgtacacag ttattagtc caccagctcc aacggaagaa gatgttatga 1620
agctcgttaag cgtcgttacc caattgctt ctttagttcc accagatcgt caagctgctt 1680
taataggtga tttattcattc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740
tggccgcaga gaatcgttta cagcaaaaaa agagttagtt ggaaggaagg actgaagtga 1800
accatgctaa tacaatgaa gaagttccct ccaggcgaac aagaagttaga gacacaatg 1860
caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacga 1920
aaaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgtaatact tctagggtat 1980
gatccaatat caaaggaaat gatagcattt aaggatgaga ctaatccaat tgaggagtgg 2040
cagcatatag aacagctaaa gggtagtgtt gaaggaagca tacgataccc cgcatggaat 2100
gggataatat cacaggaggt actagactac ctttcatttc acataaatacg acgcatataa 2160
gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220
caacacgcag atatagggtc gacgtgaaca gtgagctgtt tgcgcgcagc tcgcgttgca 2280
tttcggaaag cgctcgaaaa cggaaacgct ttgaagttcc tattccgaag ttcctattct 2340
ctagaaagta taggaacttc agagcgctt tgaaaaccaa aagcgctctg aagacgcact 2400
ttcaaaaaaac caaaaacgca ccggactgtt acgagctact aaaatattgc gaataccgct 2460
tccacaaaca ttgctaaaaa gatatcttt gctatatac tctgtgttat atcccttat 2520
aacctaccca tccaccttc gtccttcaa cttgcattta aactcgacct ctacatcaac 2580
aggcttccaa tgcttttcaa attttactgt caagtagacc catacggttg taatatgttg 2640
ctcttcataa tgtaagctt tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700
ctttacggtt ccctgagatt gaatttagttc ctttagtata tgatacaaga cactttgaa 2760
ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaattt tcctgtcggc 2820

gaaacatgct gcttaaaaact ccaagcggta ggagaccgat aaaggtaat aggacagccg 2880
tattatctcc gcctcagttt gatcttccgc ttcagactgc cattttcac ataatgaatc 2940
tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 3000
atgttgtaca ttgttttagtt cacgagaagg gtccttca ggcggtagct cctgatctcc 3060
tatatgacct ttatcctgtt ctcttccac aaacttagaa atgtattcat gaattatgga 3120
gcacctaata acattctca aggcggagaa gtttggcca gatgccaaat atgcttgaca 3180
tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3240
tcgaagataa gagaagaatg cagtgaccc ttgtattgaca aatggagatt ccatgtatct 3300
aaaaaaatacg cctttaggcc ttctgatacc ctccccctg cggttagcg tgccctttac 3360
attaatatct aaaccctctc cgatggtggc cttaactga ctaataaatg caaccgatat 3420
aaactgtgat aattctgggt gatttatgat tcgatcgaca attgtattgt acactagtgc 3480
aggatcaggc caatccagtt cttttcaat taccggtgtg tcgtctgtat tcagtagatg 3540
tccaacaaat gcaaattgcta acgttttgc tttcttataa ttgtcaggaa ctggaaaagt 3600
cccccttgtc gtctcgatta cacacctact ttcatcgta accatagggtt ggaagtgctg 3660
cataatacat tgcttaatac aagcaagcag tctctcgcca ttcatatttc agttatttc 3720
cattacagct gatgtcattt gatattcagcg ctgtaaaaat ctatctgtt cagaaggttt 3780
tcgcggtttt tataaacaaa actttcgta cgaaatcgag caatcccc agctgcgtat 3840
ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcata 3900
ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaac ctcaatctgt 3960
ccattgaatg ccttataaaa cagctataga ttgcataagaa gagttgcta ctcaatgctt 4020
tttgcattttt cttactgtatg atgatgtgct tactttcagg cgggtctgtt gtaaggagaa 4080
tgacattata aagctggcac ttagaattcc acggactata gactataacta gtataactccg 4140
tctactgtac gatacacttc cgctcaggctc cttgtccctt aacgaggcct taccactctt 4200
ttgttactct attgatccag ctcagcaaag gcagtgtat ctaagattct atcttcgcga 4260
tgttagaaaa ctagctagac cgagaaaagag actagaaaatg caaaaggcac ttctacaatg 4320
gctgccatca ttattatccg atgtgacgct gcattttttt tttttttttt tttttttttt 4380
tttttttttt tttttttttt tttttttgtt caaatatcat aaaaaaaagag aatctttta 4440
agcaaggatt ttcttaactt cttcggcgac agcatcaccg acttcggtg 4500
accacctaata tcaccaggatc tgatacctgc atccaaaacc ttttttaactg catcttcaat 4560
ggctttaccc tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4620

agtggcgata gggttgacct tattcttgg caaatctgga gcggaaaccat ggcatggttc 4680
gtacaaacca aatgcggtgt tcttgctgg caaagaggcc aaggacgcag atggcaacaa 4740
acccaaggag cctggataa cgaggcgttc atcgagatg atatcaccaa acatgttgct 4800
ggtgattata ataccattta ggtgggttgg gttcttaact agatcatgg cgccagaatc 4860
aatcaattga tgtgaactt tcaatgttagg gaattcggtt ttgtgggtt cctccacagt 4920
tttctccat aatcttgaag aggccaaaac attagctta tccaaggacc aaataggcaa 4980
tggggctca tgttgtaggg ccatgaaagc ggccattctt gtgattctt gcacttctgg 5040
aacggtgtat tgttcaactat cccaagcgac accatcacca tcgtcttcct ttctcttacc 5100
aaagtaataa cctcccacta attctctaac aacaacgaag tcagtagctt tagcaaattg 5160
tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5220
ggcgtacaat tgaagttctt tacggattt tagtaaacct tggtcaggc taacactacc 5280
ggtacccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttgaggc 5340
ttccagcgcc tcatctggaa gtggAACACC tgtagcatcg atagcagcac caccaattaa 5400
atgattttcg aaatcgaact tgacatttgg acgaacatca gaaatagctt taagaacctt 5460
aatggcttcg gctgtgattt cttgaccaac gtggcacctt ggcaaaaacga cgatcttctt 5520
aggggcagac attacaatgg tatatccttgg aaatatataaaaaaaa aaaaaaaaaa 5580
aaaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5640
tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5700
acatccgaac ctgggagttt tccctgaaac agatagtata tttgaacctt tataataata 5760
tatagtcttag cgcttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5820
atctattgca tagtaatct tgcacgtcgc atccccgggtt cattttctgc gtttccatct 5880
tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcatttttgg tagaacaaaaa 5940
atgcaacgcg agagcgctaa ttttcaaac aaagaatctg agctgcattt ttacagaaca 6000
gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tggcttcattttgtaaaa 6060
caaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca 6120
gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttcttttttg 6180
ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatcttta gattactttt 6240
tttctccctt gtgcgtcta taatgcagtc tcttgataac ttttgcact gtaggtccgt 6300
taaggttaga agaaggctac tttgggtgtct attttcttccctt ccataaaaaaa agcctgactc 6360
cacttccgc gtttactgat tactagcgaa gctgcgggtg catttttca agataaaggc 6420

atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtata 6480
gcgttcatga ttcttcattt gtcagaaaat tatgaacggt ttcttctatt ttgtcttat 6540
atactacgta taggaaatgt ttacatttc gtattgttt cgattcaactc tatgaatagt 6600
tcttactaca attttttgt ctaaagagta atactagaga taaacataaa aaatgttagag 6660
gtcgaggta gatgcaagtt caaggagcga aagggtggatg ggtaggat ataggatat 6720
agcacagaga tatatacgaa agagatactt ttgagcaatg tttgtgaaag cggtattcgc 6780
aatatttttag tagctcgta cagtcgggtc cgtttttgtt ttttgaaag tgcgtttca 6840
gagcgcttt ggtttcaaa agcgctctga agttcctata ctttctagag aataggaact 6900
tcggaatagg aacctcaaag cgttccgaa aacgagcgct tccgaaaatg caacgcgagc 6960
tgcgcacata cagctcaactg ttcacgtcgc acctatatct gcgtgttgc tgtatata 7020
tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7080
atttatgttag gatgaaaggt agtctagttc ctcctgtat attatcccattccatgcggg 7140
gtatcgtatg ctcccttcag cactaccctt tagctgtctt atatgctgcc actcctcaat 7200
tggatttagtc tcatccttca atgctatcat ttcccttgat attggatcat atgcataatgta 7260
ccgagaaact agtgcgaagt agtgcgttgcggtt atctgtatgag tatacgttgc 7320
cctggccacg gcagaagcac gcttattcgct ccaatttcccc acaacattag tcaactccgt 7380
taggccttc attgaaagaa atgagggtcat caaatgtctt ccaatgtgag attttggcc 7440
atttttata gcaaagattt gaaaaaggcgc atttttcttc aaagctttat tgtacgtct 7500
gactaagtta tctttataa attggatttc ctgtttatttgc cttgaagaat tgccggctt 7560
atttactcgt tttaggactg gttcagaatt cctcaaaaat tcataccaaat atacaagtgg 7620
atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7680
tattttata ggttaatgtc atgataataa tggttctta gacgtcagggt ggcacttttc 7740
ggggaaatgt gcgcggacc cctatttttttatttctta aatacattca aatatgtatc 7800
cgctcatgag acaataaccc tgataaaatgc ttcaataata ttgaaaaagg aagagtatga 7860
gtattcaaca ttccgtgtc gcccttatttgc cttttttgc ggcattttgc cttccgttt 7920
ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7980
tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagttt cgccccgaag 8040
aacgtttcc aatgatgagc acttttaaag ttctgctatg tggcgccgtt ttatcccgtt 8100
ttgacgcccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttt 8160
agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 8220

tgctgccat aaccatgagt gataaacactg cggccaaactt acttctgaca acgatcgag 8280
gaccgaagga gctaaccgct ttttgcaca acatggggga tcgttaact cgccctgatc 8340
gttggaaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8400
cagcaatggc aacaacgttg cgcaaaactat taactggcga actacttact cttagttccc 8460
ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccacctt ctgcgctcgg 8520
cccttcggc tggctggttt attgctgata aatctggagc cggtgagcgt gggctcgcg 8580
gtatcattgc agcaactgggg ccagatggta agccctcccg tatacgtagtt atctacacga 8640
cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8700
tgattaagca ttggtaactg tcagaccaag tttactcata tatacttttag attgatttaa 8760
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgcataat ctcatgacca 8820
aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8880
gatcttcttgc agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 8940
cgctaccagc ggtgggttgtt ttgcggatc aagagctacc aactctttt ccgaaggtaa 9000
ctggcttcag cagagcgcag ataccaaata ctgtccttct agttagccg tagttaggcc 9060
accacttcaa gaactctgta gcacccgcata catacctcgc tctgctaattc ctgttaccag 9120
tggctgctgc cagtgccgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9180
cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9240
gaacgaccta caccgaactg agataacctac agcgtgagct atgagaaagc gccacgcctc 9300
ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcgaaaca ggagagcgca 9360
cgagggagct tccagggggaa aacgccttgtt atctttatag tccctgtcggg tttcgccacc 9420
tctgacttga gcgtcgattt ttgtgatgtc cgtcagggggg gggggccta tggaaaaacg 9480
ccagcaacgc ggcctttta cgggtcctgg cctttgctg gcctttgtt cacatgttct 9540
ttcctgcgtt atccctgtat tctgtggata accgttattac cgcctttgag tgagctgata 9600
ccgctcgccg cagccgaacg accgagcgcga gcgagtcagt gagcgaggaa gcggaaagagc 9660
gcctgatgcg gtatttctc cttacgcattc tgtcggtat ttcacaccgc atatggtgca 9720
ctctcagttac aatctgctct gatgccgtat agttaagcca gtatacactc cgctatcgct 9780
acgtgactgg gtcgtggctg cggccggaca cccgccaaca cccgctgacg cggccctgacg 9840
ggcttgcgtc ctcggcat ccgttacag acaagctgtc accgtctccg ggagctgcat 9900
gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctgcggtaa agctcattc 9960
agcgtggcgtc tgaagcgatt cacagatgtc tgcctgttca tccgcgttcca gtcgtttag 10020

tttctccaga agcgtaatg tctggttct gataaagcgg gccatgttaa gggcggttt 10080
ttcctgttg gtcactgatg cctccgtgta agggggattt ctgttcatgg gggtaatgat 10140
accgatgaaa cgagagagga tgctcacat acgggttact gatgatgaac atgcccgggt 10200
actggaacgt tgtgaggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10260
caactcagggt caatgccagc gcttcgttaa tacagatgta ggtgttccac agggtagcca 10320
gcagcatcct gcgatgcaga tccggaacat aatggtgcag ggcgctgact tccgcgttcc 10380
cagactttac gaaacacgga aaccgaagac cattcatgat gttgctcagg tcgcagacgt 10440
tttgcagcag cagtcgcgttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10500
aaggcaaccc cgccagccta gccgggtctt caacgacagg agcacgatca tgcgcacccg 10560
tggccaggac ccaacgctgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10620
gatggatatg ttctgccaag ggttggtttgc cgcattcaca gttctccgca agaattgatt 10680
ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggc 10740
gaggtggccc ggctccatgc accgcgacgc aacgcggggaa ggcagacaag gtatagggcg 10800
gcccctacaa tccatgcca cccgttccat gtgctgcggc aggcggcata aatgcgcgtg 10860
acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10920
tgtccctgat ggtcgtcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10980
atgcccggc aagcgagaag aatcataatg gggaaaggcca tccagcctcg cgtcgcgaac 11040
gccagcaaga cgtagccccag cgctcgccgc gccatgcggc cgataatggc ctgcttcg 11100
ccgaaacgtt tggggcggg accagtgcg aaggcttgag cgagggcgtg caagattccg 11160
aataccgcaa gcgacaggcc gatcatcgatc ggcgtccagc gaaagcggc ctcgcccggaa 11220
atgaccggaga gcgctgcggc cacctgtcct acgagttgca tgataaagaa gacagtata 11280
agtgcggcga cgatagtcat gccccgcgc caccggaaagg agctgactgg gttgaaggct 11340
ctcaaggggca tcggtcgagg atccctcaat atgcgcacat acgctgttat gttcaaggc 11400
ccttcgttta agaacgaaag cggcttcct tttgaggat gttcaagtt gttcaatct 11460
atcaaatttg caaatccccca gtctgtatct agagcggtga atcggtatg cgattgtta 11520
attaaattga tgggtgtcacc attaccaggat ctagatatac caatggcaaa ctgagcaca 11580
caataccaggat cggatcaac tggcaccatc tctccgttag tctcatctaa ttttcttcc 11640
ggatgagggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11700
atgtcccatt cgaaatcacc aattctaaac ctgggcgaat tgtatccgg gtttgttaac 11760
tcgttccagt caggaatgtt ccacgtgaag ctatcttcca gcaaagtctc cacttcttca 11820

tcaaatttgt gagaatactc ccaatgctct tatctatggg acttccggga aacacagtac 11880
 cgatacttcc caattcgtct tcagagctca ttgtttgtt gaagagacta atcaaagaat 11940
 cgtttctca aaaaaattaa tatcttaact gatagttga tcaaaggggc aaaacgtagg 12000
 ggcaaacaaa cgaaaaaatc gtttctcaaa tttctgtatg ccaagaactc taaccagtct 12060
 tatctaaaaa ttgccttatg atccgtctct ccgggtacag cctgtgtaac tgattaatcc 12120
 tgccttcta atcaccattc taatgtttta attaaggat tttgtttca ttaacggctt 12180
 tcgctcataa aaatgttatg acgtttgcc cgccaggccc aaaccatcca ctgcacgaga 12240
 ctgatctcct ctgccggAAC accgggcatac tccaaacttat aagttggaga aataagagaa 12300
 tttcagattt agagaatgaa aaaaaaaaaac ccttagttca taggtccatt ctcttagcgc 12360
 aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtgtatg 12420
 caacctgcct ggagtaaatg atgacacaaag gcaattgacc cacgcacgtatcat 12480
 tttcttacac cttcttattac cttctgtct ctctgatttg gaaaaagctg aaaaaaaaaagg 12540
 ttgaaaccag ttccctgaaa ttattccct acttgactaa taagtatata aagacggtag 12600
 gtattgattt taattctgtatcatc aatctatttc taaaacttct taaattctac ttttatagtt 12660
 agtctttttt ttagtttaa aacaccaaga acttagtttc gaataaaacac acataaacaa 12720
 acaagcttac aaaacaaatt cacc atg gct gca tat gca gct cag ggc tat 12771
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr
 1 5
 aag gtg cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt 12819
 Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly
 10 15 20 25
 gct tac atg tcc aag gct cat ggg atc gat cct aac atc agg acc ggg 12867
 Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly
 30 35 40
 gtg aga aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc 12915
 Val Arg Thr Ile Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly
 45 50 55
 aag ttc ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac ata ata 12963
 Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile
 60 65 70
 att tgt gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc att 13011
 Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile
 75 80 85
 ggc act gtc ctt gac caa gca gag act gcg ggg ggc aga ctg gtt gtg 13059
 Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val
 90 95 100 105
 ctc gcc acc gcc acc cct ccg ggc tcc gtc act gtg ccc cat ccc aac 13107

Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn			
110	115	120	
atc gag gag gtt gct ctg tcc acc acc gga gag atc cct ttt tac ggc	13155		
Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly			
125	130	135	
aag gct atc ccc ctc gaa gta atc aag ggg ggg aga cat ctc atc ttc	13203		
Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe			
140	145	150	
tgt cat tca aag aag tgc gac gaa ctc gcc gca aag ctg gtc gca	13251		
Cys His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala			
155	160	165	
ttg ggc atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg tcc gtc	13299		
Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val			
170	175	180	185
atc ccg acc agc ggc gat gtt gtc gtc gtg gca acc gat gcc ctc atg	13347		
Ile Pro Thr Ser Gly Asp Val Val Val Ala Thr Asp Ala Leu Met			
190	195	200	
acc ggc tat acc ggc gac ttc gac tcg gtg ata gac tgc aat acg tgt	13395		
Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys			
205	210	215	
gtc acc cag aca gtc gat ttc agc ctt gac cct acc ttc acc att gag	13443		
Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu			
220	225	230	
aca atc acg ctc ccc caa gat gct gtc tcc cgc act caa cgt cgg ggc	13491		
Thr Ile Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly			
235	240	245	
agg act ggc agg ggg aag cca ggc atc tac aga ttt gtg gca ccg ggg	13539		
Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly			
250	255	260	265
gag cgc ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat	13587		
Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr			
270	275	280	
gac gca ggc tgt gct tgg tat gag ctc acg ccc gcc gag act aca gtt	13635		
Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val			
285	290	295	
agg cta cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac	13683		
Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp			
300	305	310	
cat ctt gaa ttt tgg gag ggc gtc ttt aca ggc ctc act cat ata gat	13731		
His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp			
315	320	325	
gcc cac ttt cta tcc cag aca aag cag agt ggg gag aac ctt cct tac	13779		
Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr			
330	335	340	345

ctg gta gcg tac caa gcc acc gtg tgc gct agg gct caa gcc cct ccc Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro 350 355 360	13827
cca tcg tgg gac cag atg tgg aag tgt ttg att cgc ctc aag ccc acc Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr 365 370 375	13875
ctc cat ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt cag aat Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn 380 385 390	13923
gaa atc acc ctg acg cac cca gtc acc aaa tac atc atg aca tgc atg Glu Ile Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met 395 400 405	13971
tcg gcc gac ctg gag gtc gtc acg agc acc tgg gtc ctc gtt ggc ggc Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly 410 415 420 425	14019
gtc ctg gct ttg gcc gcg tat tgc ctg tca aca ggc tgc gtc gtc Val Leu Ala Ala Leu Ala Tyr Cys Leu Ser Thr Gly Cys Val Val 430 435 440	14067
ata gtg ggc agg gtc gtc ttg tcc ggg aag ccg gca atc ata cct gac Ile Val Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp 445 450 455	14115
agg gaa gtc ctc tac cga gag ttc gat gag atg gaa gag tgc tct cag Arg Glu Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln 460 465 470	14163
cac tta ccg tac atc gag caa ggg atg atg ctc gcc gag cag ttc aag His Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys 475 480 485	14211
cag aag gcc ctc ggc ctc ctg cag acc gcg tcc cgt cag gca gag gtt Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val 490 495 500 505	14259
atc gcc cct gtc cag acc aac tgg caa aaa ctc gag acc ttc tgg Ile Ala Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp 510 515 520	14307
gcg aag cat atg tgg aac ttc atc agt ggg ata caa tac ttg gcg ggc Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly 525 530 535	14355
ttg tca acg ctg cct ggt aac ccc gcc att gct tca ttg atg gct ttt Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe 540 545 550	14403
aca gct gct gtc acc agc cca cta acc act agc caa acc ctc ctc ttc Thr Ala Ala Val Thr Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe 555 560 565	14451
aac ata ttg ggg ggg tgg gtg gct gcc cag ctc gcc gcc ccc ggt gcc Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala 570 575 580 585	14499

gct act gcc ttt gtg ggc gtc ggc tta gct ggc gcc gcc atc ggc agt Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser 590	595	600	14547	
gtt gga ctg ggg aag gtc ctc ata gac atc ctt gca ggg tat ggc gcg Val Gly Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala 605	610	615	14595	
ggc gtg gcg gga gct ctt gtg gca ttc aag atc atg agc ggt gag gtc Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val 620	625	630	14643	
ccc tcc acg gag gac ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro 635	640	645	14691	
gga gcc ctc gta gtc ggc gtg gtc tgt gca gca ata ctg cgc cgg cac Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His 650	655	660	665	14739
gtt ggc ccg ggc gag ggg gca gtg cag tgg atg aac cgg ctg ata gcc Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala 670	675	680	14787	
ttc gcc tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg cgg gag Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu 685	690	695	14835	
agc gat gca gct gcc cgc gtc act gcc ata ctc agc agc ctc act gta Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val 700	705	710	14883	
acc cag ctc ctg agg cga ctg cac cag tgg ata agc tcg gag tgt acc Thr Gln Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr 715	720	725	14931	
act cca tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg ata tgc Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys 730	735	740	745	14979
gag gtg ttg agc gac ttt aag acc tgg cta aaa gct aag ctc atg cca Glu Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro 750	755	760	15027	
cag ctg cct ggg atc ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg Gln Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly 765	770	775	15075	
gtc tgg cga ggg gac ggc atc atg cac act cgc tgc cac tgt gga gct Val Trp Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala 780	785	790	15123	
gag atc act gga cat gtc aaa aac ggg acg atg agg atc gtc ggt cct Glu Ile Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro 795	800	805	15171	
agg acc tgc agg aac atg tgg agt ggg acc ttc ccc att aat gcc tac Arg Thr Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr 810	815	820	15219	

acc acg ggc ccc tgt acc ccc ctt cct gcg ccg aac tac acg ttc gcg Thr Thr Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala 830 835 840	15267
cta tgg agg gtg tct gca gag gaa tac gtg gag ata agg cag gtg ggg Leu Trp Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly 845 850 855	15315
gac ttc cac tac gtg acg ggt atg act act gac aat ctt aaa tgc ccg Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Leu Lys Cys Pro 860 865 870	15363
tgc cag gtc cca tcg ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc Cys Gln Val Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg 875 880 885	15411
cta cat agg ttt gcg ccc ccc tgc aag ccc ttg ctg cgg gag gag gta Leu His Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val 890 895 900 905	15459
tca ttc aga gta gga ctc cac gaa tac ccg gta ggg tcg caa tta cct Ser Phe Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro 910 915 920	15507
tgc gag ccc gaa ccg gac gtg gcc gtg ttg acg tcc atg ctc act gat Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp 925 930 935	15555
ccc tcc cat ata aca gca gag gcg gcc ggg cga agg ttg gcg agg gga Pro Ser His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly 940 945 950	15603
tca ccc ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca Ser Pro Pro Ser Val Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro 955 960 965	15651
tct ctc aag gca act tgc acc gct aac cat gac tcc cct gat gct gag Ser Leu Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu 970 975 980 985	15699
ctc ata gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile 990 995 1000	15747
acc agg gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat Thr Arg Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp 1005 1010 1015	15795
ccg ctt gtg gcg gag gag gac gag cgg gag atc tcc gta ccc gca gaa Pro Leu Val Ala Glu Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu 1020 1025 1030	15843
atc ctg cgg aag tct cgg aga ttc gcc cag gcc ctg ccc gtt tgg gcg Ile Leu Arg Lys Ser Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala 1035 1040 1045	15891
cgg ccg gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac Arg Pro Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp 1050 1055 1060 1065	15939

tac gaa cca cct gtg gtc cat ggc tgc ccc ctt cca cct cca aag tcc Tyr Glu Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser 1070 1075 1080	15987
cct cct gtg cct ccc cgg aag aag cgg acg gtg gtc ctc act gaa Pro Pro Val Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu 1085 1090 1095	16035
tca acc cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc Ser Thr Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly 1100 1105 1110	16083
agc tcc tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct Ser Ser Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Ser Ser 1115 1120 1125	16131
gag ccc gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat Glu Pro Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr 1130 1135 1140 1145	16179
tcc tcc atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser 1150 1155 1160	16227
gac ggg tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc Asp Gly Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val 1165 1170 1175	16275
gtg tgc tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg Val Cys Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro 1180 1185 1190	16323
tgc gcc gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg Cys Ala Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser 1195 1200 1205	16371
ttg cta cgt cac cac aat ttg gtg tat tcc acc acc tca cgc agt gct Leu Leu Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala 1210 1215 1220 1225	16419
tgc caa agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac Cys Gln Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp 1230 1235 1240	16467
agc cat tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa Ser His Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ser Lys 1245 1250 1255	16515
gtg aag gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc Val Lys Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro 1260 1265 1270	16563
cca cac tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt Pro His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg 1275 1280 1285	16611
tgc cat gcc aga aag gcc gta acc cac aac tcc gtg tgg aaa gac Cys His Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp 1290 1295 1300 1305	16659

ctt ctg gaa gac aat gta aca cca ata gac act acc atc atg gct aag 16707
 Leu Leu Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys
 1310 1315 1320

aac gag gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct 16755
 Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala
 1325 1330 1335

cgt ctc atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg 16803
 Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met
 1340 1345 1350

gct ttg tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc 16851
 Ala Leu Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser
 1355 1360 1365

tcc tac gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg 16899
 Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val
 1370 1375 1380 1385

caa gcg tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc 16947
 Gln Ala Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr
 1390 1395 1400

cgc tgc ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag 16995
 Arg Cys Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu
 1405 1410 1415

gca atc tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc 17043
 Ala Ile Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile
 1420 1425 1430

aag tcc ctc acc gag agg ctt tat gtt ggg ggc cct ctt acc aat tca 17091
 Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser
 1435 1440 1445

agg ggg gag aac tgc ggc tat cgc agg tgc cgc gcg agc ggc gta ctg 17139
 Arg Gly Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu
 1450 1455 1460 1465

aca act agc tgt ggt aac acc ctc act tgc tac atc aag gcc cgg gca 17187
 Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala
 1470 1475 1480

gcc tgt cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc 17235
 Ala Cys Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly
 1485 1490 1495

gac gac tta gtc gtt atc tgt gaa agc gcg ggg gtc cag gag gac gcg 17283
 Asp Asp Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala
 1500 1505 1510

gcg agc ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc 17331
 Ala Ser Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro
 1515 1520 1525

cct ggg gac ccc cca caa cca gaa tac gac ttg gag ctc ata aca tca 17379
 Pro Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser
 1530 1535 1540 1545

tgc tcc tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc Cys Ser Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val 1550 1555 1560	17427
tac tac ctc acc cgt gac cct aca acc ccc ctc gcg aga gct gcg tgg Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp 1565 1570 1575	17475
gag aca gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc Glu Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile 1580 1585 1590	17523
atg ttt gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc Met Phe Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe 1595 1600 1605	17571
ttt agc gtc ctt ata gcc agg gac cag ctt gaa cag gcc ctc gat tgc Phe Ser Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys 1610 1615 1620 1625	17619
gag atc tac ggg gcc tgc tac tcc ata gaa cca ctg gat cta cct cca Glu Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro 1630 1635 1640	17667
atc att caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac Ile Ile Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr 1645 1650 1655	17715
tct cca ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg Ser Pro Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly 1660 1665 1670	17763
gta ccg ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct Val Pro Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala 1675 1680 1685	17811
agg ctt ctg gcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc Arg Leu Leu Ala Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu 1690 1695 1700 1705	17859
ttc aac tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala 1710 1715 1720	17907
gct ggc cag ctg gac ttg tcc ggc tgg ttc acg gct ggc tac agc ggg Ala Gly Gln Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly 1725 1730 1735	17955
gga gac att tat cac agc gtg tct cat gcc cgg ccc cgc tgg atc tgg Gly Asp Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp 1740 1745 1750	18003
ttt tgc cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ctc ccc Phe Cys Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro 1755 1760 1765	18051
aac cga tgaaggttgg ggtaaacact ccggcctaaa aaaaaaaaaa aatctagaac Asn Arg 1770	18107

ccgagtcgac tttgttccca ctgtactttt agctcgata aaataacaata tactttcat 18167
ttctccgtaa acaacatgtt ttcccatgta atatcctttt ctatTTTcg ttccgttacc 18227
aactttacac atacttata tagctattca cttctataca ctaaaaaact aagacaattt 18287
taatTTTgct gcctgccata tttcaatttg ttataaattc ctataatttta tcctattagt 18347
agctaaaaaa agatgaatgt gaatcgaatc ctaagagaat tggatctgat ccacaggacg 18407
ggtgtggtcg ccatgatcgc gtagtcgata gtggctccaa gtagcgaagc gaggcaggact 18467
gggcggcggc caaagcggc ggacagtgtc ccgagaacgg gtgcgcatacg aaattgcatac 18527
aacgcatacata gcgcatacg cacgcatacg tgactggcga tgctgtcgga atggacgata 18587
tcccgcaga ggccggcag tacccgcata accaagccta tgcctacagc atccagggtg 18647
acgggtgccga ggatgacgat gagcgcattt ttagatttca tacacgggtc ctgactgcgt 18707
tagcaatttta actgtgataa actaccgcattaa taaagctttt tctttccaat tttttttttt 18767
tcgtcattat aaaaatcatt acgaccgaga ttcccggtta ataactgata taattaaattt 18827
gaagctctaa ttgtgagtt tagtatacat gcattttactt ataatacagt ttttagttt 18887
tgctggccgc atcttctcaa atatgcttcc cagcctgcattt ttctgttaacg ttcaccctct 18947
accttagcat cccttccctt tgcaaatagt cctcttccaa caataataat gtcagatcct 19007
gtagagacca catcatccac ggttctatac tggtagccca atgcgtctcc cttgtcatct 19067
aaacccacac cgggtgtcat aatcaaccaa tcgtaacctt catctttcc acccatgtct 19127
ctttgagcaa taaagccgat aacaaaatct ttgtcgctct tcgcaatgtc aacagtaccc 19187
ttagtatatt ctccagtaga tagggagccc ttgcatacgat attctgttaa catcaaaagg 19247
cctcttagttt cctttgttac ttcttctgcc gcctgcttca aaccgcttaac aatacctggg 19307
cccaccacac cggtgtgcatt cgtaatgtct gcccattctg ctattctgtta tacacccgca 19367
gagtaactgca atttgactgt attaccaatg tcagcaaattt ttctgtcttc gaagagtaaa 19427
aaattgtact tggcgataa tgcctttagc ggcttaactg tgccctccat ggaaaaatca 19487
gtcaagatat ccacatgtgt ttttagtaaa caaattttgg gacctaattgc ttcaactaac 19547
tccagtaatt ccttgggtgtt acgaacatcc aatgaaggcac acaagtttgtt ttgctttcg 19607
tgcataat taaatagctt ggcagcaaca ggactaggat gagtagcagc acgttcctta 19667
tatgttagctt tcgacatgtat ttatcttcgt ttccctgcagg tttttgttct gtgcagttgg 19727
gttaagaata ctggcaattt tcatgtttct tcaacactac atatgcgtat atataccat 19787
ctaagtctgt gtccttcct tcgttctcc ttctgttcgg agattaccga atcaaaaaaa 19847
tttcaaggaa accgaaatca aaaaaaaagaa taaaaaaaaa atgatgaattt gaaaagctta 19907

tcgat 19912

<210> 9
<211> 1771
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: pd.deltaNS3NS5

<400> 9
Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser
65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys
145 150 155 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala
165 170 175

Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
180 185 190

Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe
195 200 205

Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe
210 215 220

Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp
225 230 235 240

Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro
245 250 255

Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe
 260 265 270

Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr
 275 280 285

Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn
 290 295 300

Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly
 305 310 315 320

Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr
 325 330 335

Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr
 340 345 350

Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp
 355 360 365

Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu
 370 375 380

Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro
 385 390 395 400

Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val
 405 410 415

Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala
 420 425 430

Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu
 435 440 445

Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu
 450 455 460

Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln
 465 470 475 480

Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu
 485 490 495

Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
 500 505 510

Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe
 515 520 525

Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
 530 535 540

Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
 545 550 555 560

Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575

Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590

Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605

Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
 610 615 620

Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
 625 630 635 640

Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655

Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
 660 665 670

Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
 675 680 685

Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
 690 695 700

Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
 705 710 715 720

His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735

Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750

Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765

Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
 770 775 780

Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
 785 790 795 800

Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
 805 810 815

Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
 820 825 830

Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
 835 840 845

Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
 850 855 860

Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
 865 870 875 880

Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
 885 890 895

Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910

Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
 915 920 925

Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
 930 935 940

Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
 945 950 955 960

Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975

Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
 980 985 990

Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
 995 1000 1005

Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
 1010 1015 1020

Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
 1025 1030 1035 1040

Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
 1045 1050 1055

Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070

Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
 1075 1080 1085

Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
 1090 1095 1100

Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
 1095 1110 1115 1120

Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135

Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
 1140 1145 1150

Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
 1155 1160 1165

Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
 1170 1175 1180

Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 1185 1190 1195 1200

Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
 1205 1210 1215

Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
 1220 1225 1230

Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
 1235 1240 1245

Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
 1250 1255 1260

Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 1265 1270 1275 1280

Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
 1285 1290 1295

Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310

Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325

Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 1330 1335 1340

Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
 345 1350 1355 1360

Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
 1365 1370 1375

Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
 1380 1385 1390

Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
 1395 1400 1405

Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
 1410 1415 1420

Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
 425 1430 1435 1440

Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
 1445 1450 1455

Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470

Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
 1475 1480 1485

Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
 1490 1495 1500

Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
 505 1510 1515 1520

Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535

Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
 1540 1545 1550

Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
 1555 1560 1565

Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
 1570 1575 1580

Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
 585 1590 1595 1600

Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
 1605 1610 1615

Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
 1620 1625 1630

Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
 1635 1640 1645

Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
 1650 1655 1660

Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
 665 1670 1675 1680

Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
 1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser
 1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
 1730 1735 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala
 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1765 1770

<210> 10
 <211> 19798
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 pd.deltANS3NS5.pj

<220>
 <221> CDS
 <222> (12679)..(17991)

<400> 10
atcgatccta ccccttgcgc taaaagaagta tatgtgccta ctaacgcttg tctttgtctc 60
tgtcaactaaa cactggatta ttactcccag atacttattt tggactaatt taaatgattt 120
cgatatcaacg ttcttaatat cgctgaatct tccacaattt atgaaagttag cttagaaagag 180
gaattggtat aaagttttt 400
tttctcagtg atctcccaga tgcttcacc ctcacttaga agtgctttaa gcatttttt 300
actgtggcta tttcccttat ctgcttcttc cgatgattcg aactgtaatt gcaaactact 360
tacaatataca gtgatatacg attgatgttt ttgtccatag taaggaaataa ttgtaaattc 420
ccaaggcagga atcaatttct ttaatgaggc ttccagaatt gttgctttt gcgtcttgta 480
tttaaactgg agtgatttat tgacaatatac gaaactcagc gaattgctta tgatagtatt 540
atagctcatg aatgtggctc tcttgattgc ttgtccgtta tgtgtaatca tccaacataa 600
ataggtagt tcagcagcac ataatgctat tttctcacct gaaggcttt caaaccttcc 660
cacaactga cgaacaagca ccttaggtgg ttgtttacat aatataatcaa attgtggcat 720
gcttagcgcc gatcttgcgt gcaattgata tctagttca actactctat ttatcttgta 780
tcttgagta ttcaaacacg ctaactcgaa aaactaactt taattgtcct gttgtctcg 840
cggtatcttc atttcatatt ttaaaaatgc acctttgctg ctttcctta attttttagac 900
ggccccgagg ttctttgc ggtactatct tttgtataaaa agttgtttt acatgtgatc 1020
tgcacagatt ttataatgta ataagcaaga atacattatc aaacgaacaa tactggtaaa 1080
agaaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140
cttatacggtt ctggatgta tttttttttt gttttttttt gttttttttt gttttttttt 1200
tttgatata gatgtttttttt gttttttttt gttttttttt gttttttttt gttttttttt 1260
ccatgaaatc tctcacaacc ggtggccgt atggcttga ttctagcgca gaagattcca 1320
gcgtatcttc tgactccagt gctgaggtaa ttttgcctgc tgcgaagatg gtttggaaa 1380
ggtttggattc gatgtttttttt gttttttttt gttttttttt gttttttttt gttttttttt 1440
atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattt 1500
ctataataat aggaagattt cccgagaaaag acaagaagag agctaccgaa atgctcatga 1560
gaaaaatgga ttgtacacag ttattgtcc caccagctcc aacgaaagaaa gatgttatga 1620
agctcgtaag cgtcgatacc caattgctta ctttagttcc accagatcgta caagctgctt 1680
taataggtga ttatccatc ccggaaatctc taaaggatat attcaatagt ttcaatgaac 1740

tggcggcaga gaatcgttt cagcaaaaa agagttagtt ggaaggaaagg actgaagtga 1800
accatgctaa tacaaatgaa gaagttccct ccaggcgaac aagaagtaga gacacaatg 1860
caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttccacga 1920
aaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgttaatact tctaggtat 1980
gatccaatat caaaggaaat gatagcattt aaggatgaga ctaatccaat tgaggagtgg 2040
cagcatatag aacagctaaa gggttagtgct gaaggaagca tacgatacccg cgcatggaat 2100
gggataatat cacaggaggt actagactac ctttcatcct acataaaatag acgcataataa 2160
gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220
caacacgcag atatagggtc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280
tttcggaaag cgctcgaaaa cgaaaaacgct ttgaagttcc tattccgaag ttcctattct 2340
ctagaaaagta taggaacttc agagcgctt taaaacccaa aagcgctctg aagacgcact 2400
ttcaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460
tccacaaaca ttgctcaaaa gatatcttt gctatataatc tctgtgctat atccctatata 2520
aacctaccca tccacccccc gtccttgaa cttgcataa aactcgacct ctacatcaac 2580
aggcttccaa tgctctcaa attttactgt caagtagacc catacggtctg taatatgctg 2640
ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700
ctttacggtt ccctgagatt gaatttagtt cttagtata tgataacaaga cactttgaa 2760
ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaattt tcctgtcggc 2820
tattatctcc gcctcagttt gatctccgc ttcagactgc cattttcac ataatgaatc 2880
tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940
atgttgatac ttgttagtt cacgagaagg gtcctttca ggcggtagct cctgatctcc 3000
tatatgaccc ttatcctgtt ctcttcacaa aacttagaa atgtattcat gaattatgga 3060
gcacctaata acattctca aggccggagaa gtttggcca gatgcccata atgcttgaca 3120
tggaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180
tcgaagataa gagaagaatg cagtgaccc ttgtattgaca aatggagatt ccatgtatct 3240
aaaaaaatcg cctttaggcc ttctgatacc ctccccctg cggttagcg tgcctttac 3300
attaatatct aaaccctctc cgatggtggc cttaactga ctaataaaatg caaccgatata 3360
aaactgtgat aattctgggt gatttatgat tcgatcgaca attgtattgt acactagtgc 3420
aggatcagggc caatccagtt cttttcaat taccgggtgtg tcgtctgtat tcagtacatg 3480
tccaacaaat gcaaatgcta acgttttgta tttcttataa ttgtcaggaa ctggaaaagt 3540

cccccttgcgtc gtctcgatta cacacctaact ttcatcgta accataggtt ggaagtgctg 3600
cataatacat tgcttaataac aagcaagcag tctctcgcca ttcatatttc agttatccc 3660
cattacagct gatgtcattt tatatcagcg ctgtaaaaat ctatctgtta cagaaggaaa 3720
tcgcggtttt tataaacaaa actttcgta cgaaatcgag caatcacccc agctgcgtat 3780
ttggaaattc gggaaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcata 3840
ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaac ctcaatctgt 3900
ccattgaatg cctataaaaa cagctataga ttgcataaaaaa gagtttagcta ctcaatgctt 3960
tttgtcaaaag cttactgtatg atgatgtgtc tacttcagg cgggtctgtt gtaaggagaa 4020
tgacattata aagctggcac tttagaattcc acggactata gactatacta gtataactccg 4080
tctactgtac gatacacttc cgctcaggc tttgtccctt aacgaggcct taccactttt 4140
ttgttactctt attgatccag ctcagcaaag gcagtgtatg ctaagattctt atttcgcga 4200
tgttagaaaaa ctagctagac cgagaaagag actagaaatg caaaaggcac ttctacaatg 4260
gctgccatca ttattatccg atgtgacgct gcattttttt tttttttttt tttttttttt 4320
ttttttttttt tttttttttt tttttttggta caaatatcat aaaaaaaagag aatctttta 4380
agcaaggatt ttcttaactt cttcggcgac agcatcacccg acttcgggtgg tactgttgaa 4440
accacctaaa tcaccaggatc tgataacctgc atccaaaacc tttttaactg catcttcaat 4500
ggctttaccc tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560
agtggcgata gggttgaccc tattttttgg caaatctggc gcggaaaccat ggcattggc 4620
gtacaaacca aatgcgggtgt tcttgcgtgg caaagaggcc aaggacgcag atggcaacaa 4680
acccaaaggag cctgggataa cggaggcttc atcggagatg atatcacccaa acatgttgct 4740
ggtgattata ataccattta ggtgggttgg gttcttaactt aggtatcatgg cggcagaatc 4800
aatcaattga tgttgaaactt tcaatgttagg gaattcggtt ttgtatggttt cctccacagt 4860
ttttctccat aatcttgaag aggcacaaac attagcttta tccaggacc aaataggcaa 4920
tgggtggctca tgttgtaggg ccatgaaagc ggccattttt gtgattctt gcacttctgg 4980
aacgggttat tggactat cccaaagcgc accatcacca tcgtttccct ttcttccacc 5040
aaagtaataa cctcccaacta attctctaacc aacaacgaag tcagtacccctt tagcaaattg 5100
tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5160
ggcgtacaat tgaagttctt tacggatttt tagtaaacct tggatggc taacactacc 5220
ggtagcccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttgaggc 5280
ttccagcgcc tcatactggaa gtggaaacacc tggatgcatacg atagcagcacc caccaattaa 5340

atgatttcg aaatcgaact tgacattgga acgaacatca gaaatagctt taagaacctt 5400
aatggcttcg gctgtgattt cttgaccaac gtggcacct ggcaaaacga cgatttctt 5460
aggggcagac attacaatgg tatatccttg aaatatataat aaaaaaaaaaaa aaaaaaaaaaaa 5520
aaaaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5580
tatccgacaa actgtttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640
acatccgaac ctgggagttt tccctgaaac agatagtata tttgaacctg tataataata 5700
tatagtcttag cgcttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5760
atcttattgca taggtaatct tgcacgtcgc atccccggtt cattttctgc gttccatct 5820
tgcaactcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaaaa 5880
atgcaacgcg agagcgctaa ttttcaaac aaagaatctg agctgcattt ttacagaaca 5940
gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat tttttaaaaa 6000
caaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca 6060
gaacagaaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttctttttt 6120
ttctacaaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta gattactttt 6180
tttctccctt gtgcgtctta taatgcagtc tcttgataac ttttgcact gtaggtccgt 6240
taaggtaga agaaggctac tttggtgtct attttctctt ccataaaaaaa agcctgactc 6300
cactccccgc gtttactgat tactagcga gctgcgggtg catttttca agataaaggc 6360
atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6420
gcgttcatga ttcttcattt gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6480
atactacgta taggaaatgt ttacatttc gtattgtttt cgattcactc tatgaatagt 6540
tcttactaca attttttgcataaagataatgttataaataaa aatgttagag 6600
gtcgagttt gatgcaagtt caaggagcga aagggtggatg ggttaggttat ataggatata 6660
agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaaag cggtattcgc 6720
aatatttttag tagctcgatca cagtcgggtg cgttttgggt ttttgaag tgcgtcttca 6780
gagcgctttt ggtttcaaa agcgctctga agttcctata ctttctagag aataggaact 6840
tcggaatagg aacttcaaag cgttccgaa aacgagcgct tccgaaaatg caacgcgagc 6900
tgcgcacata cagctcactg ttcacgtcgc acctatatct gcgtgttgcc tgtatata 6960
tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7020
atttatgttag gatgaaaggt agtcttagtac ctcctgtat attatcccatttccatgcggg 7080
gtatcgtatg cttccattcag cactaccctt tagctgttct atatgcgtgcc actcctcaat 7140

tggattagtc tcatccttca atgcttatcat ttcccttgat attggatcat atgcatacgta 7200
ccgagaaact agtgcgaagt agtgatcagg tattgctgtt atctgatgag tatacgttgt 7260
cctggccacg gcagaagcac gcttatacgct ccaatttccc acaacattag tcaactccgt 7320
taggcccttc attgaaaagaa atgaggtcat caaatgtctt ccaatgtgag attttggcc 7380
atttttata gcaaagattg aataaggcgc attttcttc aaagctttat tgtacgatct 7440
gactaagtta tctttaata attggatttc ctgttattg cttgaagaat tgccggcct 7500
atttactcgt ttaggactg gttcagaatt cctaaaaat tcatccaaat atacaagtgg 7560
atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7620
atttttata ggttaatgtc atgataataa tggttctta gacgtcagg ggcactttc 7680
ggggaaatgt gcgcggaacc cctattgtt tattttcta aatacattca aatatgtatc 7740
cgctcatgag acaataaccc tgataaaatgc ttcaataata ttgaaaaagg aagagtatga 7800
gtattcaaca ttccgtgtc gcccatttc ccttttgc ggcattttgc cttccgttt 7860
ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7920
tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagttt cgccccgaag 7980
aacgtttcc aatgatgagc actttaaag ttctgctatg tggcgcggtt ttatcccgtg 8040
ttgacgcccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 8100
agtaactcacc agtcacagaaa aagcatcttta cgatggcat gacagtaaga gaattatgca 8160
gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcgag 8220
gaccgaagga gctaaccgct ttttgacaca acatggggta tcatgttaact cgccttgatc 8280
gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340
cagcaatggc aacaacgttg cgcaaactat taactggcgactacttact ctagcttccc 8400
ggcaacaatt aatagactgg atggaggcgg ataaaagttgc aggaccactt ctgcgctcgg 8460
cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg 8520
gtatcattgc agcaactgggg ccagatggta agccctcccg tatcgttagtt atctacacga 8580
cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640
tgattaagca ttggtaactg tcagaccaag tttactcata tatacttttag attgatttaa 8700
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttataat ctcatgacca 8760
aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820
gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 8880
cgctaccagc ggtggtttgtt ttgccggatc aagagctacc aactctttt ccgaaggtaa 8940

ctggcttcag cagagcgcag ataccaaata ctgtccttct agttagccg tagttaggcc 9000
accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaatt ctgttaccag 9060
tggctgtgc cagtgccat aagtctgtc ttaccgggtt ggactcaaga cgatagttac 9120
cgatataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180
gaacgaccta caccgaactg agataacctac agcgtgagct atgagaaaagc gccacgcttc 9240
ccgaaggggag aaaggcggac aggtatccgg taagcggcag gtcggaaaca ggagagcgca 9300
cgagggagct tccagggggaa aacgccttgtt atctttatag tcctgtcggg tttcgccacc 9360
tctgacttga gcgtcgattt ttgtgtatgtc cgtcaggggg gcggagccta tgaaaaacg 9420
ccagcaacgc ggcctttta cggttcctgg cctttgtc gcctttgtc cacatgttct 9480
ttcctgcgtt atccccgtat tctgtggata accgtattac cgccttgag tgagctgata 9540
ccgctcgcccg cagccgaacg accgagcgcga gcgagtcagt gagcggaggaa gcggaaagagc 9600
gcctgatgcg gtatttctc cttacgcattc tgtgcgttat ttcacaccgc atatggtgca 9660
ctctcagttac aatctgtctc gatgccgcattt agttaagccaa gtatacactc cgctatcgct 9720
acgtgactgg gtcattggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780
ggcttgcgtc ctcccgcat ccgttacag acaagctgtc accgtctccg ggagctgcatt 9840
gtgtcagagg ttttccaccgt catcaccgaa acgcgcgagg cagctgcgtt aaagctcatc 9900
agcgtggtcg tgaagcgatt cacagatgtc tgcctgttca tccgcgtcca gtcgtttag 9960
tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020
ttcctgttttgcgtactgtatg cctccgttca agggggattt ctgttcatgg ggtaatgtat 10080
accgatgaaa cgagagagga tgctcacat acgggttact gatgtgaac atgcccggtt 10140
actggaaacgt tggatggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200
caactcagggt caatgccagc gcttcgttaa tacagatgtc ggtgttccac aggtagccaa 10260
gcagcatcct gcgtgcaga tccggaacat aatggtgcag ggctgtact tccgcgtttc 10320
cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380
tttgcagcag cagtcgcttc acgttcgttc gcgtatcggt gattcattct gctaaccagt 10440
aaggcaaccc cgccagccta gcccgggttca acacgacagg acacgatca tgccgcacccg 10500
tggccaggac ccaacgctgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10560
gatggatatg ttctgccaag ggttggtttc cgcattcaca gttctccgca agaattgatt 10620
ggctccaatt ctggatgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggc 10680
gaggtggccc ggctccatgc accgcgcacgc aacgcggggaa ggcagacaag gtatagggcg 10740

gcgcctacaa tccatgccaa cccgttccat gtgctgccg aggcggcata aatcccgtg 10800
acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10860
tgtccctgat ggtcgcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10920
atgccgcccgg aagcgagaag aatcataatg gggaggcca tccagcctcg cgtcgcaac 10980
gccagcaaga cgtagcccag cgcgtcgccg gccatgccgg cgataatggc ctgcttcg 11040
ccgaaacgtt tggggcggg accagtgacg aaggcttgag cgagggcgtg caagattccg 11100
aataccgcaa gcgacaggcc gatcatcgtc ggcgtccagc gaaagcggtc ctcgcccggaa 11160
atgaccaga gcgcgtccgg cacctgtcct acgagttgca tgataaagaa gacagtata 11220
agtgcggcga cgatagtcat gccccgcgc caccggagg agctgactgg gttgaaggct 11280
ctcaaggcga tcggtcgagg atccttcaat atgcgcacat acgctgttat gttcaaggc 11340
ccttcgttta agaacgaaaag cggcttcct tttgagggat gttcaagtt gttcaaatct 11400
atcaaatttg caaatccccca gtctgtatct agagcgttga atcggtgatg cgatttgtta 11460
attaaattga tgggtgcacc attaccaggc cttagatatac caatggcaaa ctgagcacaa 11520
caataccagt ccggatcaac tggcaccatc tctccgtag tctcatctaa ttttcttcc 11580
ggatgagggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11640
atgtcccatt cgaaatcacc aattctaaac ctggcgaat tgtatccgg gtttgttaac 11700
tcgttccagt caggaatgtt ccacgtgaag ctatcttca gcaaagtctc cacttcttca 11760
tcaaattgtg gagaataactc ccaatgctct tatctatggg acttccggga aacacagtac 11820
cgatacttcc caattcgctc tcagagctca ttgtttgttt gaagagacta atcaaagaat 11880
cgtttctca aaaaaattaa tatcttaact gatagttga tcaaaggggc aaaacgtagg 11940
ggcaaacaaa cggaaaaatc gtttctcaaa ttttctgatg ccaagaactc taaccagtct 12000
tatctaaaaa ttgccttatg atccgtctct ccggttacag cctgtgttaac tgattaatcc 12060
tgccttctca atcaccatc taatgtttta attaagggat ttgttcttca ttaacggctt 12120
tcgctcataa aaatgttatg acgtttgcc cgcaggcggg aaaccatcca cttcacgaga 12180
ctgatctcct ctgcccgaac accgggcac tccaaacttat aagttggaga aataagagaa 12240
tttcagattg agagaatgaa aaaaaaaaaac ctttagttca taggtccatt ctcttagcgc 12300
aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtgtatg 12360
caacctgcct ggagtaaatg atgacacaag gcaattgacc cacgcatgtt tctatctcat 12420
tttcttacac cttcttattac cttctgtctc ctctgatttg gaaaaagctg aaaaaaaaaagg 12480
ttgaaaccag ttccctgaaa ttattccct acttgactaa taagtatata aagacggtag 12540

gtattgattttaattctgtataatctatttc ttAAacttctaaattctac ttttatagtt 12600
 agtctttttt ttagtttaaacaccaaga acttagttc gaataaacac acataaacaa 12660
 acaagcttac aaaacaaa atg gct gca tat gca gct cag ggc tat aag gtg 12711
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val
 1 5 10
 cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac 12759
 Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr
 15 20 25
 atg tcc aag gct cat ggg atc gat cct aac atc agg acc ggg gtg aga 12807
 Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg
 30 35 40
 aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc 12855
 Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe
 45 50 55
 ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt 12903
 Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys
 60 65 70 75
 gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc att ggc act 12951
 Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr
 80 85 90
 gtc ctt gac caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc gcc 12999
 Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala
 95 100 105
 acc gcc acc cct ccg ggc tcc gtc act gtg ccc cat ccc aac atc gag 13047
 Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu
 110 115 120
 gag gtt gct ctg tcc acc acc gga gag atc cct ttt tac ggc aag gct 13095
 Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala
 125 130 135
 atc ccc ctc gaa gta atc aag ggg ggg aga cat ctc atc ttc tgt cat 13143
 Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His
 140 145 150 155
 tca aag aag aag tgc gac gaa ctc gcc gca aag ctg gtc gca ttg ggc 13191
 Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly
 160 165 170
 atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg tcc gtc atc ccg 13239
 Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro
 175 180 185
 acc agc ggc gat gtt gtc gtc gtg gca acc gat gcc ctc atg acc ggc 13287
 Thr Ser Gly Asp Val Val Val Ala Thr Asp Ala Leu Met Thr Gly
 190 195 200
 tat acc ggc gac ttc gac tcg gtg ata gac tgc aat acg tgt gtc acc 13335
 Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr
 205 210 215

cag aca gtc gat ttc agc ctt gac cct acc ttc acc att gag aca atc Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile 220 225 230 235	13383
acg ctc ccc caa gat gct gtc tcc cgc actcaa cgt cgg ggc agg act Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr 240 245 250	13431
ggc agg ggg aag cca ggc atc tac aga ttt gtg gca ccg ggg gag cgc Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg 255 260 265	13479
ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gca Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala 270 275 280	13527
ggc tgt gct tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu 285 290 295	13575
cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat ctt Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu 300 305 310 315	13623
gaa ttt tgg gag ggc gtc ttt aca ggc ctc act cat ata gat gcc cac Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His 320 325 330	13671
ttt cta tcc cag aca aag cag agt ggg gag aac ctt cct tac ctg gta Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val 335 340 345	13719
gcg tac caa gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser 350 355 360	13767
tgg gac cag atg tgg aag tgt ttg att cgc ctc aag ccc acc ctc cat Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His 365 370 375	13815
ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa atc Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile 380 385 390 395	13863
acc ctg acg cac cca gtc acc aaa tac atc atg aca tgc atg tcg gcc Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala 400 405 410	13911
gac ctg gag gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu 415 420 425	13959
gct gct ttg gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata gtg Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val 430 435 440	14007
ggc agg gtc gtc ttg tcc ggg aag ccg gca atc ata cct gac agg gaa Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu 445 450 455	14055

gtc ctc tac cga gag ttc gat gag atg gaa gag tgc tct cag cac tta		14103	
Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu			
460	465	470	475
ccg tac atc gag caa ggg atg atg ctc gcc gag cag ttc aag cag aag		14151	
Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys			
480	485	490	
gcc ctc ggc ctc ctg cag acc gcg tcc cgt cag gca gag gtt atc gcc		14199	
Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala			
495	500	505	
cct gct gtc cag acc aac tgg caa aaa ctc gag acc ttc tgg gcg aag		14247	
Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys			
510	515	520	
cat atg tgg aac ttc atc agt ggg ata caa tac ttg gcg ggc ttg tca		14295	
His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser			
525	530	535	
acg ctg cct ggt aac ccc gcc att gct tca ttg atg gct ttt aca gct		14343	
Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala			
540	545	550	555
gct gtc acc agc cca cta acc act agc caa acc ctc ctc ttc aac ata		14391	
Ala Val Thr Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile			
560	565	570	
ttg ggg ggg tgg gtg gct gcc cag ctc gcc ccc ggt gcc gct act		14439	
Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr			
575	580	585	
gcc ttt gtg ggc gct ggc tta gct ggc gcc gcc atc ggc agt gtt gga		14487	
Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly			
590	595	600	
ctg ggg aag gtc ctc ata gac atc ctt gca ggg tat ggc gcg ggc gtg		14535	
Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val			
605	610	615	
gcg gga gct ctt gtg gca ttc aag atc atg agc ggt gag gtc ccc tcc		14583	
Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser			
620	625	630	635
acg gag gac ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc gga gcc		14631	
Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala			
640	645	650	
ctc gta gtc ggc gtg gtc tgt gca gca ata ctg cgc cgg cac gtt ggc		14679	
Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly			
655	660	665	
ccg ggc gag ggg gca gtg cag tgg atg aac cgg ctg ata gcc ttc gcc		14727	
Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala			
670	675	680	
tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat		14775	
Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp			
685	690	695	

gca gct gcc cgc gtc act gcc ata ctc agc agc ctc act gta acc cag Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln 700 705 710 715	14823
ctc ctg agg cga ctg cac cag tgg ata agc tcg gag tgt acc act cca Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro 720 725 730	14871
tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val 735 740 745	14919
ttg agc gac ttt aag acc tgg cta aaa gct aag ctc atg cca cag ctg Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu 750 755 760	14967
cct ggg atc ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp 765 770 775	15015
cga ggg gac ggc atc atg cac act cgc tgc cac tgt gga gct gag atc Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile 780 785 790 795	15063
act gga cat gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr 800 805 810	15111
tgc agg aac atg tgg agt ggg acc ttc ccc att aat gcc tac acc acg Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr 815 820 825	15159
ggc ccc tgt acc ccc ctt cct gcg ccg aac tac acg ttc gcg cta tgg Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp 830 835 840	15207
agg gtg tct gca gag gaa tac gtg gag ata agg cag gtg ggg gac ttc Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe 845 850 855	15255
cac tac gtg acg ggt atg act act gac aat ctt aaa tgc ccg tgc cag His Tyr Val Thr Gly Met Thr Asp Asn Leu Lys Cys Pro Cys Gln 860 865 870 875	15303
gtc cca tcg ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cat Val Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His 880 885 890	15351
agg ttt gcg ccc ccc tgc aag ccc ttg ctg cgg gag gag gta tca ttc Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe 895 900 905	15399
aga gta gga ctc cac gaa tac ccg gta ggg tcg caa tta cct tgc gag Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu 910 915 920	15447
ccc gaa ccg gac gtg gcc gtg ttg acg tcc atg ctc act gat ccc tcc Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser 925 930 935	15495

cat ata aca gca gag gcg gcc ggg cga agg ttg gcg agg gga tca ccc His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro 940 945 950 955	15543
ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc Pro Ser Val Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu 960 965 970	15591
aag gca act tgc acc gct aac cat gac tcc cct gat gct gag ctc ata Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile 975 980 985	15639
gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg 990 995 1000	15687
gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu 1005 1010 1015	15735
gtg gcg gag gag gac gag cg ^g gag atc tcc gta ccc gca gaa atc ctg. Val Ala Glu Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu 1020 1025 1030 1035	15783
cg ^g aag tct cgg aga ttc gcc cag gcc ctg ccc gtt tgg gcg cg ^g ccg Arg Lys Ser Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro 1040 1045 1050	15831
gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu 1055 1060 1065	15879
cca cct gtg gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro 1070 1075 1080	15927
gtg cct ccg cct ccg aag aag cgg acg gtg gtc ctc act gaa tca acc Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr 1085 1090 1095	15975
cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser 1100 1105 1110 1115	16023
tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro 1120 1125 1130	16071
gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser 1135 1140 1145	16119
atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly 1150 1155 1160	16167
tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys 1165 1170 1175	16215

tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala 1180 1185 1190 1195	16263
gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu 1200 1205 1210	16311
cgt cac cac aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln 1215 1220 1225	16359
agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His 1230 1235 1240	16407
tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ser Lys Val Lys 1245 1250 1255	16455
gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His 1260 1265 1270 1275	16503
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His 1280 1285 1290	16551
gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu 1295 1300 1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1310 1315 1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1325 1330 1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1340 1345 1350 1355	16743
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr 1360 1365 1370	16791
gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala 1375 1380 1385	16839
tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1390 1395 1400	16887
ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca atc Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile 1405 1410 1415	16935

tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc		16983
Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser		
1420 1425 1430 1435		
ctc acc gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg		17031
Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly		
1440 1445 1450		
gag aac tgc ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act		17079
Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr		
1455 1460 1465		
agc tgt ggt aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt		17127
Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys		
1470 1475 1480		
cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac		17175
Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp		
1485 1490 1495		
tta gtc gtt atc tgt gaa agc gcg ggg gtc cag gag gac gca gca gac		17223
Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser		
1500 1505 1510 1515		
ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg		17271
Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly		
1520 1525 1530		
gac ccc cca caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc		17319
Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser		
1535 1540 1545		
tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac		17367
Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr		
1550 1555 1560		
ctc acc cgt gac cct aca acc ccc ctc gcg aga gct gca tgg gag aca		17415
Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr		
1565 1570 1575		
gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt		17463
Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe		
1580 1585 1590 1595		
gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc		17511
Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser		
1600 1605 1610		
gtc ctt ata gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc		17559
Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile		
1615 1620 1625		
tac ggg gcc tgc tac tcc ata gaa cca ctg gat cta cct cca atc att		17607
Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile		
1630 1635 1640		
caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca		17655
Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro		
1645 1650 1655		

ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg 17703
 Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro
 1660 1665 1670 1675

 ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt 17751
 Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu
 1680 1685 1690

 ctg gcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac 17799
 Leu Ala Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn
 1695 1700 1705

 tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc 17847
 Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly
 1710 1715 1720

 cag ctg gac ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac 17895
 Gln Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp
 1725 1730 1735

 att tat cac agc gtg tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc 17943
 Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys
 1740 1745 1750 1755

 cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga 17991
 Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1760 1765 1770

 tgaatagtgcg actttgttcc cactgtactt tttagctcgta caaaaatacaa tatactttc 18051
 atttctccgt aaacaacatg ttttcccatg taatatcctt ttctattttt cgttccgtta 18111
 ccaactttac acatacttta tatagttatt cacttctata cactaaaaaa ctaagacaat 18171
 tttaattttg ctgcctgcca tatttcaatt tgttataaat tcctataatt tatttttattt 18231
 gtagctaaaa aaagatgaat gtgaatcgaa tcctaaagaga attggatctg atccacagga 18291
 cgggtgtggc cgccatgatc gcgttagtcga tagtggtc aagtagcgaa gcgagcagga 18351
 ctggggcgccg gccaaagcgg tcggacagtgc tccgagaac gggtgcgcatt agaaattgc 18411
 tcaacgcata tagcgctagc agcacgcatt agtgactggc gatgctgtcg gaatggacga 18471
 tatcccccaa gagggccggc agtaccggca taaccaagcc tatgcctaca gcatccaggg 18531
 tgacggtgcc gagggatgacg atgagcgcatt tgtagattt catacacggc gcctgactgc 18591
 gtttagcaatt taactgtgat aaactaccgc attaaagctt ttcttttcca atttttttt 18651
 ttctgtcatt ataaaaatca ttacgaccga gattcccgaa taataactga tataattaaa 18711
 ttgaagctct aatttgtgag tttagtatac atgcattac ttataataca gtttttttagt 18771
 tttgctggcc gcatcttctc aaatatgctt cccagcctgc ttttctgtaa cgttcaccct 18831
 ctaccttagc atcccttccc tttgcaaata gtcctttcc aacaataata atgtcagatc 18891
 ctgttagagac cacatcatcc acgggttctat actgttgacc caatgcgtct cccttgcatt 18951

ctaaaaccac accgggtgtc ataatcaacc aatcgtaacc ttcatctctt ccacccatgt 19011
ctcttgagc aataaagccg ataacaaaat ctttgcgtc ctgcgaatg tcaacagtac 19071
ccttagtata ttctccagta gatagggagc cttgcattga caattctgct aacatcaaaa 19131
ggcctctagg ttcccttgtt acttcttcgt ccgcctgctt caaaccgcta acaataacctg 19191
ggcccaccac accgtgtgca ttcgtaatgt ctgcccattc tgctattctg tatacacccg 19251
cagagtactg caatttgact gtattaccaa tgtcagcaaa ttttctgtct tcgaagagta 19311
aaaaattgtt cttggcggat aatgccttta gcggcttaac tgtgccctcc atggaaaaat 19371
cagtcaagat atccacatgt gtttttagta aacaaatttt gggacctaatt gttcaacta 19431
actccagtaa ttccctggtg gtacgaacat ccaatgaagc acacaagttt gttgctttt 19491
cgtgcattgt attaaatagc ttggcagcaa caggactagg atgagtagca gcacgttcc 19551
tatatgttagc ttgcacatg atttatcttc gtttcctgca gtttttgtt ctgtgcagtt 19611
gggttaagaa tactggcua tttcatgttt cttcaacact acatatgcgt atatatacca 19671
atctaagtct gtgccttc cttcggtt cttctgttc ggagattacc gaatcaaaaa 19731
aatttcaagg aaaccgaaat caaaaaaaaaaag aaaaaaaaaa aatgatgaa ttgaaaagct 19791
tatcgat 19798

<210> 11
<211> 1771
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
pd.deltaNS3NS5.pj

<400> 11
Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
 50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys Cys His Ser
65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
 100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
 115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
 130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys
 145 150 155 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala
 165 170 175

Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
 180 185 190

Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe
 195 200 205

Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe
 210 215 220

Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp
 225 230 235 240

Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro
 245 250 255

Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe
 260 265 270

Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr
 275 280 285

Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn
 290 295 300

Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly
 305 310 315 320

Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr
 325 330 335

Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr
 340 345 350

Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp
 355 360 365

Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu
 370 375 380

Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro
 385 390 395 400

Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val
 405 410 415

Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala
 420 425 430

 Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu
 435 440 445

 Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu
 450 455 460

 Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln
 465 470 475 480

 Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu
 485 490 495

 Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
 500 505 510

 Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe
 515 520 525

 Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
 530 535 540

 Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
 545 550 555 560

 Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575

 Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590

 Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605

 Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
 610 615 620

 Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
 625 630 635 640

 Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655

 Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
 660 665 670

 Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
 675 680 685

 Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
 690 695 700

 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
 705 710 715 720

 His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735

Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750

Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765

Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
 770 775 780

Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
 785 790 795 800

Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
 805 810 815

Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
 820 825 830

Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
 835 840 845

Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
 850 855 860

Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
 865 870 875 880

Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
 885 890 895

Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910

Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
 915 920 925

Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
 930 935 940

Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
 945 950 955 960

Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975

Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
 980 985 990

Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
 995 1000 1005

Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
 1010 1015 1020

Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
 1025 1030 1035 1040

Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
 1045 1050 1055

Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070
 Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
 1075 1080 1085
 Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
 1090 1095 1100
 Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
 105 1110 1115 1120
 Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135
 Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
 1140 1145 1150
 Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
 1155 1160 1165
 Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
 1170 1175 1180
 Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 1185 1190 1195 1200
 Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
 1205 1210 1215
 Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
 1220 1225 1230
 Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
 1235 1240 1245
 Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
 1250 1255 1260
 Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 1265 1270 1275 1280
 Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
 1285 1290 1295
 Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310
 Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325
 Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 1330 1335 1340
 Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
 1345 1350 1355 1360
 Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
 1365 1370 1375

Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
 1380 1385 1390

Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
 1395 1400 1405

Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
 1410 1415 1420

Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
 425 1430 1435 1440

Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
 1445 1450 1455

Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470

Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
 1475 1480 1485

Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
 1490 1495 1500

Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
 505 1510 1515 1520

Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535

Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
 1540 1545 1550

Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
 1555 1560 1565

Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
 1570 1575 1580

Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
 585 1590 1595 1600

Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
 1605 1610 1615

Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
 1620 1625 1630

Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
 1635 1640 1645

Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
 1650 1655 1660

Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
 665 1670 1675 1680

Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
 1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser
 1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
 1730 1735 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Leu Ala
 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1765 1770

<210> 12

<211> 20220

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core121

<220>

<221> CDS

<222> (12679)..(18354)

<400> 12

atcgatccta ccccttgcgc taaagaagta tatgtgccta ctaacgcttg tctttgtctc 60
 tgtcaactaaa cactggatta ttactcccag atacttattt tggactaatt taaatgattt 120
 cgatatcaacg ttcttaatat cgctgaatct tccacaattt atgaaagttag cttagaagag 180
 gaattggtat aaagtttttg ttttgtaaa tctcgaagta tactcaaacg aatttagtat 240
 tttctcagtg atctcccaga tgctttcacc ctcacttaga agtgctttaa gcatttttt 300
 actgtggcta tttcccttat ctgcttccttc cgatgattcg aactgtaatt gcaaactact 360
 tacaatatca gtgatatcag attgatgttt ttgtccatag taaggaataa ttgtaaattc 420
 ccaagcagga atcaatttct ttaatgaggc ttccagaatt gttgctttt gcgtcttgta 480
 tttaaactgg agtgatttat tgacaatatc gaaactcagc gaattgctta tgatagtatt 540
 atagctcatg aatgtggctc tcttgattgc tggccgtta tggtaatca tccaacataa 600
 ataggtagt tcagcagcac ataatgctat tttctcacct gaaggctttt caaaccttcc 660
 cacaaactga cgaacaagca ccttaggtgg tggatcat aatatacaa attgtggcat 720
 gcttagcgcc gatcttgtgt gcaattgata tctagttca actactctat ttatcttgta 780
 tcttgagta ttcaaacacg ctaactcgaa aaactaactt taattgtcct gtttgtctcg 840

cttttcataa tgtaagctt tctttatcga atcggtgaa aaactactac cgcgataaac 2700
ctttacggtt ccctgagatt gaatttagttc cttagtata tgataacaaga cactttgaa 2760
cttgcacga cgaattttga ggttcgccat cctctggcta tttccaatta tcctgtcggc 2820
tattatctcc gcctcagttt gatcttccgc ttcagactgc cattttcac ataatgaatc 2880
tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940
atgttgtaca ttgttagtt cacgagaagg gtcctttca ggcggtagct cctgatctcc 3000
tatatgacct ttatcctgtt ctctttccac aaacttagaa atgtattcat gaattatgga 3060
gcacctaata acattctca aggcggagaa gtttggcca gatgcccata atgctgaca 3120
tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180
tcgaagataa gagaagaatg cagtgcaccc ttgtattgaca aatggagatt ccatgtatct 3240
aaaaaaatacg cctttaggcc ttctgatacc ctccccctg cggtttagcg tgcctttac 3300
attaatatct aaaccctctc cgatggtggc cttaactga ctaataatg caaccgatat 3360
aaactgtgat aattctgggt gatttatgat tcgatcgaca attgtattgt acactagtgc 3420
aggatcagggc caatccagtt cttttcaat taccgggtgtg tcgtctgtat tcagtacatg 3480
tccaacaaat gcaaatgcta acgttttcta tttcttataa ttgtcaggaa ctggaaaagt 3540
cccccttgcgtc gtctcgatta cacacctact ttcatcgta accataggtt ggaagtgcgt 3600
cataatacat tgcttaatac aagcaagcag tctctcgcca ttcatatttc agttatttc 3660
cattacagct gatgtcatttgc tatatcagcg ctgtaaaaat ctatctgtta cagaaggttt 3720
tcgcggtttt tataaacaaa actttcgta cgaatcgag caatcaccac agctgcgtat 3780
ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcata 3840
ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaac ctcaatctgt 3900
ccattgaatg ccttataaaa cagctataga ttgcataaaaa gagttagcta ctcaatgctt 3960
tttgcgtttt cttactgtatg atgatgtgtc tactttcagg cgggtctgtt gtaaggagaa 4020
tgacattata aagctggcac tttagaattcc acggactata gactataacta gtataactcc 4080
tctactgtac gatacacttc cgctcaggtc cttgtcctt aacgaggcct taccacttt 4140
ttgttactct attgatccag ctcagcaaag gcagtgtat ctaagattct atcttcgcga 4200
tgttagaaaa ctagctagac cgagaaaagag actagaaaatg caaaaggcac ttctacaatg 4260
gctgccatca ttattatccg atgtgacgct gcattttttt tttttttttt tttttttttt 4320
tttttttttt tttttttttt ttttttggta caaatatcat aaaaaaaagag aatctttta 4380
agcaaggatt ttcttaactt cttcggcgac agcatcaccg acttcgggtgg tactgttggaa 4440

accacctaaa tcaccagttc tgataacctgc atccaaaacc ttttaactg catcttcaat 4500
ggctttacct tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560
agtggcgata ggggtgacct tattctttgg caaatctgga gcggaaaccat ggcatggttc 4620
gtacaaacca aatgcggtgt tcttgcgtgg caaagaggcc aaggacgcag atggcaacaa 4680
acccaaggag cctgggataa cgaggagcttc atcggagatg atatcaccaa acatgttgc 4740
ggtgattata ataccattta ggtggggttgg gttcttaact aggatcatgg cgccagaatc 4800
aatcaattga tggtaactt tcaatgttagg gaattgcgtt ttgatggttt cctccacagt 4860
ttttctccat aatcttgaag aggccaaaac attagcttta tccaaggacc aaataggcaa 4920
tggtggtca tggtaggg ccatgaaagc ggccattctt gtgattctt gcacttctgg 4980
aacgggttat tggtaactat cccaaagcgac accatcacca tcgtcttcct ttctcttacc 5040
aaagtaaata cctccccacta attctctaac aacaacgaag tcagtagctt tagcaaattg 5100
tggcttgatt ggagataagt ctAAAAGAGA gtcggatgca aagttacatg gtcttaagtt 5160
ggcgtacaat tgaagttctt tacggattt tagtaaacct tggtaggtc taacactacc 5220
ggtacccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttgaggc 5280
ttccagcgcc tcatctggaa gtggAACACC tggatcgatcg atagcagcac caccaattaa 5340
atgatttcg aaatcgaact tggatggaa acgaacatca gaaatagctt taagaacctt 5400
aatggcttcg gctgtgattt cttgaccaac gtggcacct ggcaaaaacga cgatcttctt 5460
aggggcagac attacaatgg tataccttggaa aaaaaaaaaaaaaaaa 5520
aaaaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5580
tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640
acatccgaac ctggagttt tccctgaaac agatagtata tttgaacctg tataataata 5700
tatagtcttag cgctttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5760
atctattgca tagtaatct tgcacgtcgc atccccgggtt cattttctgc gtttccatct 5820
tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaAAA 5880
atgcaacgcg agagcgctaa ttttcaaac aaagaatctg agctgcattt ttacagaaca 5940
gaaatgcaac gcaacgcgc tattttacca acgaagaatc tgtgcttcat ttttgtaaaa 6000
caaaaatgca acgcgagagc gctaattttt caaacaaga atctgagctg catttttaca 6060
gaacagaaat gcaacgcgc agcgctattt taccacaaa gaatctatac ttcttttttg 6120
ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatcttta gattactttt 6180
tttctccctt gtgcgtcta taatgcagtc tcttgataac ttttgcact gtaggtccgt 6240

taaggtaga agaaggctac tttgggtct atttctctt ccataaaaaa agcctgactc 6300
cacttccgc gttactgat tactagcgaa gctgcgggtg catttttca agataaaggc 6360
atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6420
gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6480
atactacgta taggaaatgt ttacatttc gtattgttt cgattcaactc tatgaatagt 6540
tcttactaca attttttgt ctaaagagta atactagaga taaacataaaa aaatgttagag 6600
gtcgagtttta gatgcaagtt caaggagcga aagggtggatg gtaggttat atagggatat 6660
agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtgaaag cggtattcgc 6720
aatattttag tagctcgta cagtcgggtg cgttttgggt ttttgaag tgcgtcttca 6780
gagcgctttt ggaaaaaaa agcgctctga agttcctata ctttctagag aataggaact 6840
tcggaatagg aacctcaaag cgttccgaa aacgagcgct tccgaaaatg caacgcgagc 6900
tgcgcacata cagctcaactg ttcacgtcgc acctatatct gcgtgttgcc tgtatata 6960
tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7020
atttatgttag gatgaaaggt agtctagtagc ctcctgtgat attatcccattccatgcggg 7080
gtatcgatg cttcccttcag cactaccctt tagctgttct atatgctgcc actcctcaat 7140
tggattagtc tcatttcata atgctatcat ttcccttgat attggatcat atgcata 7200
ccgagaaact agtgcgaagt agtgcgttgc tattgtgtt atctgtatg tatacgttgc 7260
cctggccacg gcagaagcac gcttacgcct ccaatttcccc acaacattag tcaactccgt 7320
taggccttc attgaaagaa atgaggtcat caaatgtctt ccaatgtgag attttggggcc 7380
atttttata gcaaagattg aataaggcgc attttcttc aaagctttat tgtacgatct 7440
gactaagtta tctttataata attggatttc ctgttatttgc cttgaagaat tgccggtcct 7500
atttactcgatgttttactg gttcagaatt cctcaaaaat tcattccaaat atacaagtgg 7560
atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggccctc gtgatacgcc 7620
tatTTTATA ggttaatgtc atgataataa tggttctta gacgtcagggt ggcacttttc 7680
ggggaaatgt gcgcggAACCC CCTATTGTT TATTTCTA AATACATTCA AATATGTATC 7740
cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 7800
gtattcaaca ttccgtgtc gcccttatttgc cttttttgc ggcattttgc cttccgttt 7860
ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7920
tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagttt cgccccgaag 7980
aacgtttcc aatgatgagc acttttaaag ttctgctatg tggcgccgtt ttatcccgtg 8040

ttgacgcgg gcaagagcaa ctcggcgcc gcatacacta ttctcagaat gacttggttg 8100
agtactcacc agtcacagaa aagcatctta cgatggcat gacagtaaga gaattatgca 8160
gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcgag 8220
gaccgaagga gctaaccgct ttttgacaca acatggggga tcatgtact cgccttgatc 8280
gttggaaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340
cagcaatggc aacaacgttg cgcaaaactat taactggcga actacttact cttagttccc 8400
ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcg 8460
cccttccggc tggctggttt attgctgata aatctggagc cggtagcgt gggctcg 8520
gtatcattgc agcactgggg ccagatggta agccctcccg tatacgtagtt atctacacga 8580
cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640
tgattaagca ttgttaactg tcagaccaag tttactcata tatacttttag attgatttaa 8700
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttataat ctcatgacca 8760
aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820
gatcttcttgc agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 8880
cgctaccagc ggtgggttgt ttgcggatc aagagctacc aactctttt ccgaaggtaa 8940
ctggcttcag cagagcgcag ataccaaata ctgtccttct agttagccg tagttaggcc 9000
accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag 9060
tggctgctgc cagtgccgat aagtcgtgc ttaccgggtt ggactcaaga cgatagttac 9120
cggataaggc gcagcggcgc ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180
gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaaagc gccacgcttc 9240
ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaaca ggagagcgc 9300
cgagggagct tccagggggaa aacgcctggat atctttatag tccgtcggtt tttcgccacc 9360
tctgacttgc gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 9420
ccagcaacgc ggcctttta cggttcctgg cctttgctg gcctttgct cacatgttct 9480
ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata 9540
ccgctcgccg cagccgaacg accgagcgc ggcgtcagt gagcggagaa gcggaaagagc 9600
gcctgatgcg gtatcccgc cttacgcata tgtcggtat ttcacaccgc atatggc 9660
ctctcagttac aatctgctct gatgccgcat agttaagccaa gtatacactc cgctatcgct 9720
acgtgactgg gtcatggctg cggccgacca cccgcaaca cccgctgacg cggccgtacg 9780
ggcttgcgtt ctcggcat cgcgttacag acaagctgtg accgtctccg ggagctgcat 9840

gtgtcagagg tttcacccgt catcaccgaa acgcgcgagg cagctgcggg aaagctcatc 9900
agcgtggtcg tgaagcgatt cacagatgtc tgcctgttca tccgcgtcca gtcgttgag 9960
tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggttt 10020
ttcctgtttg gtcaactgatg cctccgtgta agggggattt ctgttcatgg gggtaatgat 10080
accgatgaaa cgagagagga tgctcacat acgggttact gatgatgaac atgcccgg 10140
actggaacgt tgtgaggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200
caactcagggt caatgccagc gcttcgttaa tacagatgta ggttccac agggtagcca 10260
gcagcatcct gcgatgcaga tccggaacat aatggtgcag ggctgact tccgcgtt 10320
cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380
tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10440
aaggcaaccc cgccagccta gcccggctt caacgacagg agcacgatca tgcgcacccg 10500
tggccaggac ccaacgctgc ccgagatgcg ccgcgtcggg ctgctggaga tggcggacgc 10560
gatggatatg ttctgccaag ggttggtttgc cgcattcaca gttctccgca agaattgatt 10620
ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggc 10680
gaggtggccc ggctccatgc accgcgacgc aacgcggggg ggcagacaag gtatagggcg 10740
gcccctacaa tccatgcca cccgttccat gtgctcggc aggcggcata aatcggcgt 10800
acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10860
tgtccctgat ggctgtcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10920
atgcgcggcgg aagcgagaag aatcataatg gggaaaggcca tccagcctcg cgtcgcgaac 10980
gccagcaaga cgtagccccag cgcgtcggcc gccatgcggg cgataatggc ctgcttcg 11040
ccgaaacgtt tggggcggg accagtgcacg aaggcttgag cgagggcgtg caagattccg 11100
aataccgcaa gcgacaggcc gatcatcgatc ggcgtccagc gaaagcggc ctcgcccggaa 11160
atgaccggaga gcgctgcggg cacctgttct acgagttgca tgataaagaa gacagtata 11220
agtgcggcga cgatagtcat gccccgcgc caccggagg agctgactgg gttgaaggct 11280
ctcaaggggca tcggtcgagg atccctcaat atgcgcacat acgctgttat gttcaaggc 11340
ccttcgttta agaacgaaag cggcttcct tttgagggat gttcaagtt gttcaatct 11400
atcaaatttg caaatccccca gtctgtatct agagcgttga atcgggtatg cgattgtta 11460
attaaattga tgggtgtcacc attaccaggat ctagatatac caatggcaaa ctgagcacaa 11520
caataccaggat cggatcaac tggcaccatc tctccgttag tctcatctaa ttttcttcc 11580
ggatgagggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11640

atgtccatt cgaaatcacc aattctaaac ctggcgaat tgtatcggtttttaac 11700
 tcgttccagt caggaatgtt ccacgtgaag ctatcttcca gcaaagtctc cacttcttca 11760
 tcaaattgtg gagaatactc ccaatgtct tatctatggg acttccggga aacacagtac 11820
 cgatacttcc caattcgtct tcagagctca ttgtttgtt gaagagacta atcaaagaat 11880
 cgaaaaattaa tatcttaact gatagttgatcaaa tttctgtatgc 11940
 ggcaaacaaaa cggaaaaatc gtttctcaaa ttgtttgtatgc 12000
 tatctaaaaaa ttgccttatg atccgtctc ccgggtacag cctgtgtac tgattaatcc 12060
 tgccttcataa atcaccattc taatgtttta attaaggat ttgttcttca ttaacggc 12120
 tcgctcataa aaatgttatg acgtttgcc cgccggggaa accatcca cttcacgaga 12180
 ctgatctcct ctgccggAAC accgggcattc tccaaacttat aagttggaga aataagagaa 12240
 tttcagattt agagaatgaa aaaaaaaaaac ctttagttca taggtccatt ctcttagcgc 12300
 aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtgtatgc 12360
 caacctgcct ggagtaaatg atgacacaag gcaattgtacc caccatgtatcat 12420
 tttcttacac cttcttattac cttctgtctc ctctgatTTT gaaaaagctg aaaaaaaaaagg 12480
 ttgaaaccag ttccctgaaa ttattccctt acttgactaa taagtatata aagacggtag 12540
 gtattgattttaa aatctatttc ttaaaacttct taaattctac ttttatagtt 12600
 agtctttttt ttagttttaa aacaccaaga acttagtttca gaataaaacac acataaaacaa 12660
 acaagcttac aaaacaaaaa atg gct gca tat gca gct cag ggc tat aag gtg 12711
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val
 1 5 10

 cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac 12759
 Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr
 15 20 25

 atg tcc aag gct cat ggg atc gat cct aac atc agg acc ggg gtg aga 12807
 Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg
 30 35 40

 aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc 12855
 Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe
 45 50 55

 ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt 12903
 Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys
 60 65 70 75

 gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc att ggc act 12951
 Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr
 80 85 90

 gtc ctt gac caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc gcc 12999

Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala			
95	100	105	
acc gcc acc cct ccg ggc tcc gtc act gtg ccc cat ccc aac atc gag	13047		
Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu			
110	115	120	
gag gtt gct ctg tcc acc acc gga gag atc cct ttt tac ggc aag gct	13095		
Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala			
125	130	135	
atc ccc ctc gaa gta atc aag ggg ggg aga cat ctc atc ttc tgt cat	13143		
Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His			
140	145	150	155
tca aag aag aag tgc gac gaa ctc gcc gca aag ctg gtc gca ttg ggc	13191		
Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly			
160	165	170	
atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg tcc gtc atc ccg	13239		
Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro			
175	180	185	
acc agc ggc gat gtt gtc gtc gtg gca acc gat gcc ctc atg acc ggc	13287		
Thr Ser Gly Asp Val Val Val Ala Thr Asp Ala Leu Met Thr Gly			
190	195	200	
tat acc ggc gac ttc gac tcg gtg ata gac tgc aat acg tgt gtc acc	13335		
Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr			
205	210	215	
cag aca gtc gat ttc agc ctt gac cct acc ttc acc att gag aca atc	13383		
Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile			
220	225	230	235
acg ctc ccc caa gat gct gtc tcc cgc act caa cgt cgg ggc agg act	13431		
Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr			
240	245	250	
ggc agg ggg aag cca ggc atc tac aga ttt gtg gca ccg ggg gag cgc	13479		
Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg			
255	260	265	
ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gca	13527		
Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala			
270	275	280	
ggc tgt gct tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta	13575		
Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu			
285	290	295	
cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat ctt	13623		
Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu			
300	305	310	315
gaa ttt tgg gag ggc gtc ttt aca ggc ctc act cat ata gat gcc cac	13671		
Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His			
320	325	330	

ttt cta tcc cag aca aag cag agt ggg gag aac ctt cct tac ctg gta Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val 335 340 345	13719
gcg tac caa gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser 350 355 360	13767
tgg gac cag atg tgg aag tgt ttg att cgc ctc aag ccc acc ctc cat Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His 365 370 375	13815
ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa atc Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile 380 385 390 395	13863
acc ctg acg cac cca gtc acc aaa tac atc atg aca tgc atg tcg gcc Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala 400 405 410	13911
gac ctg gag gtc gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu 415 420 425	13959
gct gct ttg gcc gcg tat tgc ctg tca aca ggc tgc gtc gtc ata gtg Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val 430 435 440	14007
ggc agg gtc gtc ttg tcc ggg aag ccg gca atc ata cct gac agg gaa Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu 445 450 455	14055
gtc ctc tac cga gag ttc gat gag atg gaa gag tgc tct cag cac tta Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu 460 465 470 475	14103
ccg tac atc gag caa ggg atg atg ctc gcc gag cag ttc aag cag aag Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys 480 485 490	14151
gcc ctc ggc ctc ctg cag acc gcg tcc cgt cag gca gag gtt atc gcc Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala 495 500 505	14199
cct gct gtc cag acc aac tgg caa aaa ctc gag acc ttc tgg gcg aag Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys 510 515 520	14247
cat atg tgg aac ttc att agt ggg ata caa tac ttg gcg ggc ttg tca His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser 525 530 535	14295
acg ctg cct ggt aac ccc gcc att gct tca ttg atg gct ttt aca gct Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala 540 545 550 555	14343
gct gtc acc agc cca cta acc act agc caa acc ctc ctc ttc aac ata Ala Val Thr Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile 560 565 570	14391

ttg ggg ggg tgg gtg gct gcc cag ctc gcc gcc ccc ggt gcc gct act		14439	
Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr			
575	580	585	
gcc ttt gtg ggc gct ggc tta gct ggc gcc atc ggc agt gtt gga		14487	
Ala Phe Val Gly Ala Gly Leu Ala Ala Ile Gly Ser Val Gly			
590	595	600	
ctg ggg aag gtc ctc ata gac atc ctt gca ggg tat ggc gcg ggc gtg		14535	
Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val			
605	610	615	
gcg gga gct ctt gtg gca ttc aag atc atg agc ggt gag gtc ccc tcc		14583	
Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser			
620	625	630	635
acg gag gac ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc gga gcc		14631	
Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala			
640	645	650	
ctc gta gtc ggc gtg gtc tgt gca gca ata ctg cgc cg ^g cac gtt ggc		14679	
Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly			
655	660	665	
ccg ggc gag ggg gca gtg cag tgg atg aac cgg ctg ata gcc ttc qcc		14727	
Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala			
670	675	680	
tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat		14775	
Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp			
685	690	695	
gca gct gcc cgc gtc act gcc ata ctc agc agc ctc act gta acc cag		14823	
Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln			
700	705	710	715
ctc ctg agg cga ctg cac cag tgg ata agc tcg gag tgt acc act cca		14871	
Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro			
720	725	730	
tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg		14919	
Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val			
735	740	745	
ttg agc gac ttt aag acc tgg cta aaa gct aag ctc atg cca cag ctg		14967	
Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu			
750	755	760	
cct ggg atc ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg		15015	
Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp			
765	770	775	
cga ggg gac ggc atc atg cac act cgc tgc cac tgt gga gct gag atc		15063	
Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile			
780	785	790	795
act gga cat gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc		15111	
Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr			
800	805	810	

tgc agg aac atg tgg agt ggg acc ttc ccc att aat gcc tac acc acg Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr 815	820	825	15159
ggc ccc tgt acc ccc ctt cct gcg ccg aac tac acg ttc gcg cta tgg Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp 830	835	840	15207
agg gtg tct gca gag gaa tac gtg gag ata agg cag gtg ggg gac ttc Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe 845	850	855	15255
cac tac gtg acg ggt atg act act gac aat ctt aaa tgc ccg tgc cag His Tyr Val Thr Gly Met Thr Asp Asn Leu Lys Cys Pro Cys Gln 860	865	870	15303
gtc cca tcg ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cat Val Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His 880	885	890	15351
agg ttt gcg ccc ccc tgc aag ccc ttg ctg cgg gag gag gta tca ttc Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe 895	900	905	15399
aga gta gga ctc cac gaa tac ccg gta ggg tcg caa tta cct tgc gag Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu 910	915	920	15447
ccc gaa ccg gac gtg gcc gtg ttg acg tcc atg ctc act gat ccc tcc Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser 925	930	935	15495
cat ata aca gca gag gcg gcc ggg cga agg ttg gcg agg gga tca ccc His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro 940	945	950	15543
ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc Pro Ser Val Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu 960	965	970	15591
aag gca act tgc acc gct aac cat gac tcc cct gat gct gag ctc ata Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile 975	980	985	15639
gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg 990	995	1000	15687
gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu 1005	1010	1015	15735
gtg gcg gag gag gac gag cgg gag atc tcc gta ccc gca gaa atc ctg Val Ala Glu Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu 1020	1025	1030	15783
cgx aag tct cgg aga ttc gcc cag gcc ctg ccc gtt tgg gcg cgg ccg Arg Lys Ser Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro 1040	1045	1050	15831

gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu 1055 1060 1065	15879
cca cct gtg gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro 1070 1075 1080	15927
gtg cct ccg cct ccg aag aag ccg acg gtg gtc ctc act gaa tca acc Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr 1085 1090 1095	15975
cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser 1100 1105 1110 1115	16023
tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro 1120 1125 1130	16071
gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc Ala Pro Ser Gly Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser 1135 1140 1145	16119
atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly 1150 1155 1160	16167
tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys 1165 1170 1175	16215
tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala 1180 1185 1190 1195	16263
gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu 1200 1205 1210	16311
cgt cac cac aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa Arg His His Asn Leu Val Tyr Ser Thr Ser Arg Ser Ala Cys Gln 1215 1220 1225	16359
agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His 1230 1235 1240	16407
tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ser Lys Val Lys 1245 1250 1255	16455
gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His 1260 1265 1270 1275	16503
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His 1280 1285 1290	16551

gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu 1295 1300 1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1310 1315 1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1325 1330 1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1340 1345 1350 1355	16743
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr 1360 1365 1370	16791
gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala 1375 1380 1385	16839
tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1390 1395 1400	16887
ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca atc Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile 1405 1410 1415	16935
tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser 1420 1425 1430 1435	16983
ctc acc gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly 1440 1445 1450	17031
gag aac tgc ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr 1455 1460 1465	17079
agc tgt ggt aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys 1470 1475 1480	17127
cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp 1485 1490 1495	17175
tta gtc gtt atc tgt gaa agc gcg ggg gtc cag gag gac gcg gcg agc Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser 1500 1505 1510 1515	17223
ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly 1520 1525 1530	17271

gac ccc cca caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser 1535 1540 1545	17319
tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr 1550 1555 1560	17367
ctc acc cgt gac cct aca acc ccc ctc gcg aga gct gcg tgg gag aca Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr 1565 1570 1575	17415
gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe 1580 1585 1590 1595	17463
gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser 1600 1605 1610	17511
gtc ctt ata gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile 1615 1620 1625	17559
tac ggg gcc tgc tac tcc ata gaa cca ctg gat cta cct cca atc att Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile 1630 1635 1640	17607
caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro 1645 1650 1655	17655
ggg gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro 1660 1665 1670 1675	17703
ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu 1680 1685 1690	17751
ctg gcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac Leu Ala Arg Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn 1695 1700 1705	17799
tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly 1710 1715 1720	17847
cag ctg gac ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac Gln Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp 1725 1730 1735	17895
att tat cac agc gtg tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys 1740 1745 1750 1755	17943
cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg 1760 1765 1770	17991

atg agc acg aat cct aaa cct caa aga aag acc aaa cgt aac acc aac	18039
Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn	
1775 1780 1785	
 cg ^g cg ^g cc ^g ca ^g ga ^c gtc a ^a g t ^t c cc ^g gg ^t ggc gg ^t ca ^g at ^c gtt gg ^t 18087	
Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gln Ile Val Gly	
1790 1795 1800	
 gg ^a gt ^t ta ^c tt ^g tt ^g cc ^g cg ^c ag ^g gg ^c c ^c t ag ^a tt ^g gg ^t gt ^g cg ^c gc ^g 18135	
Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala	
1805 1810 1815	
 ac ^g ag ^a a ^a g ac ^t cc ^c ga ^g cg ^g tc ^g ca ^a c ^c t ca ^g gg ^t ag ^a c ^g t ca ^g c ^c t 18183	
Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro	
1820 1825 1830 1835	
 at ^c cc ^c a ^a g g ^c t cg ^g cc ^c ga ^g gg ^c ag ^g ac ^c tt ^g g ^c t ca ^g cc ^c gg ^g 18231	
Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly	
1840 1845 1850	
 ta ^c c ^c t tt ^g cc ^c ct ^c ta ^t gg ^c aa ^a t ga ^g gg ^c t ^t g ^c gg ^g tt ^g gg ^c gg ^a tt ^g 18279	
Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp	
1855 1860 1865	
 ct ^c ct ^g t ^c t cc ^c cg ^t gg ^c t ^c t cg ^g c ^c t ag ^c tt ^g gg ^c cc ^c a ^c a g ^a c cc ^c 18327	
Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro	
1870 1875 1880	
 cg ^g cg ^t ag ^g tc ^g cg ^c a ^a t tt ^g gg ^t a ^a g taatagt ^c ga ct ^t tgttcccc 18374	
Arg Arg Arg Ser Arg Asn Leu Gly Lys	
1885 1890	
 actgtacttt tagctcgta ^c aaaatacaat atactttca tt ^t tccgt ^a aacaacatgt 18434	
tttcccatt ^g t aatacccttt tctat ^t tttc gttccgttac caactttaca catactttat 18494	
atagctattc acttctatac actaaaaaaac taagacaatt ttaat ^t ttgc tg ^c ctgccat 18554	
at ^t tcaattt gttataaatt cctataattt atcctatttag tagctaaaaaa aagatgaatg 18614	
tgaatcgaa ^t t cctaagagaa ttggatctga tccacaggac ggggtggc ^t g ^c catgatcg 18674	
cgtagtcgat agtggctcca agtagcgaag cgagcaggac tggcggcgg ccaaagcgg ^t 18734	
cggacagtgc tccgagaacg ggtgcgcata gaaattgc ^t at caacgcata ^t agcgc ^t agca 18794	
gcacgcccata gtgactggcg atgctgtcg ^t aatggacgat atcccgcaag aggcccggca 18854	
gtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg aggatgacga 18914	
tgagcgcatt gttagatttc atacacggtg cctgactgc ^t g ttagcaattt aactgtgata 18974	
aactaccgca tt ^a aagcttt ttcttccaa tttttttttt ttcgtcatta taaaaatcat 19034	
tacgaccgag attcccggt aataactgat ataattaaat tgaagctcta atttgtgagt 19094	
ttagtataca tgcatttact tataatacag ttttttagtt ttgctggccg catcttctca 19154	
aatatgcttc ccagcctgct tt ^t tctgt ^a ac gttcaccc ^c tc taccttagca tccctccct 19214	

ttgcaaatac tcctttcca acaataataa tgtcagatcc ttagagacc acatcatcca 19274
 cggttctata ctgttgaccc aatgcgtctc cttgtcatc taaacccaca cgggtgtca 19334
 taatcaacca atcgtaacct tcatctcttc caccatgtc tcttgagca ataaagccga 19394
 taacaaaatc tttgtcgctc ttcgcaatgt caacagtacc cttgtatata tctccagtag 19454
 ataggagcc cttgcattgac aattctgcta acataaaaag gcctcttagt tccttggta 19514
 cttcttcgc cgccgtctc aaaccgctaa caatacctgg gccaccaca ccgtgtgcat 19574
 tcgtaatgtc tgcccattct gctattctgt atacacccgc agagtaactgc aatttgcactg 19634
 tattaccaat gtcagcaaat tttctgtctt cgaagagtaa aaaattgtac ttggcggata 19694
 atgccttttag cggcttaact gtgccttcca tggaaaaatc agtcaagata tccacatgtg 19754
 ttttagtaa acaaattttg ggacctaattt cttcaactaa ctccagtaat tccttggtgg 19814
 tacgaacatc caatgaagca cacaagttt tttgttttc gtgcattgata taaaatagct 19874
 tggcagcaac aggacttagga tgagtagcag cacgttccctt atatgttagt ttcgacatga 19934
 tttatcttcg tttcttcgcag gttttgttc tgtcagttg gtttaagaat actgggcaat 19994
 ttcatgtttc ttcaacacta catatgcgta tatataccaa tctaagtctg tgctccttcc 20054
 ttcgttcttc cttctgttcg gagattaccg aatcaaaaaa atttcaagga aaccgaaatc 20114
 aaaaaaaaaaaga ataaaaaaaaa aatgatgaat tgaaaagctt atcgat 20160

<210> 13
 <211> 1892
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core121

<400> 13			
Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro			
1	5	10	15
Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His			
20	25	30	
Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly			
35	40	45	
Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly			
50	55	60	
Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser			
65	70	75	80
Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala			

85	90	95
Glu Thr Ala Gly Ala Arg Leu Val Val	Leu Ala Thr Ala Thr Pro Pro	
100	105	110
Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser		
115	120	125
Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val		
130	135	140
Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys		
145	150	155
Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala		
165	170	175
Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val		
180	185	190
Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe		
195	200	205
Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe		
210	215	220
Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp		
225	230	235
240		
Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro		
245	250	255
Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe		
260	265	270
Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr		
275	280	285
Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn		
290	295	300
Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly		
305	310	315
320		
Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr		
325	330	335
Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr		
340	345	350
Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp		
355	360	365
Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu		
370	375	380
Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro		
385	390	395
400		

Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val
 405 410 415
 Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala
 420 425 430
 Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu
 435 440 445
 Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu
 450 455 460
 Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln
 465 470 475 480
 Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu
 485 490 495
 Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
 500 505 510
 Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe
 515 520 525
 Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
 530 535 540
 Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
 545 550 555 560
 Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575
 Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590
 Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605
 Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
 610 615 620
 Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
 625 630 635 640
 Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655
 Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
 660 665 670
 Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
 675 680 685
 Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
 690 695 700
 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
 705 710 715 720

His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735

Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750

Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765

Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
 770 775 780

Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
 785 790 795 800

Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
 805 810 815

Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
 820 825 830

Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
 835 840 845

Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
 850 855 860

Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
 865 870 875 880

Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
 885 890 895

Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910

Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
 915 920 925

Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
 930 935 940

Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
 945 950 955 960

Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975

Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
 980 985 990

Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
 995 1000 1005

Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
 1010 1015 1020

Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
 1025 1030 1035 1040

Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
 1045 1050 1055

Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070

Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
 1075 1080 1085

Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
 1090 1095 1100

Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
 105 1110 1115 1120

Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135

Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
 1140 1145 1150

Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
 1155 1160 1165

Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
 1170 1175 1180

Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 1185 1190 1195 1200

Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
 1205 1210 1215

Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
 1220 1225 1230

Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
 1235 1240 1245

Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
 1250 1255 1260

Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 1265 1270 1275 1280

Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
 1285 1290 1295

Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310

Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325

Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 1330 1335 1340

Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
 1345 1350 1355 1360

Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
 1365 1370 1375

Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
 1380 1385 1390

Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
 1395 1400 1405

Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
 1410 1415 1420

Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
 425 1430 1435 1440

Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
 1445 1450 1455

Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470

Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
 1475 1480 1485

Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
 1490 1495 1500

Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
 505 1510 1515 1520

Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535

Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
 1540 1545 1550

Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
 1555 1560 1565

Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
 1570 1575 1580

Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
 585 1590 1595 1600

Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
 1605 1610 1615

Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
 1620 1625 1630

Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
 1635 1640 1645

Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
 1650 1655 1660

Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
 665 1670 1675 1680

Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
 1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp Leu Ser
 1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
 1730 1735 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala
 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro
 1765 1770 1775

Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
 1780 1785 1790

Val Lys Phe Pro Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu
 1795 1800 1805

Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
 1810 1815 1820

Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
 825 1830 1835 1840

Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
 1845 1850 1855

Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
 1860 1865 1870

Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
 1875 1880 1885

Asn Leu Gly Lys
 1890

<210> 14
<211> 20316
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
pd.delta.NS3NS5.pj.core173

<220>
<221> CDS
<222> (12679)..(18510)

<400> 14
atcgatccta ccccttgcgca taaagaagta tatgtgccta ctaacgcttg tctttgtctc 60

caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttccacga 1920
aaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgtataact tctaggtat 1980
gatccaatat caaaggaaat gatagcattg aaggatgaga ctaatccaat tgaggagtgg 2040
cagcatatag aacagctaaa gggtagtgct gaaggaagca tacgataccc cgcatggaat 2100
gggataatat cacaggaggt actagactac ctttcatcct acataaataag acgcatataa 2160
gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220
caacacgcag atataggtgc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280
tttcggaag cgctcgaaaa cgaaaaacgct ttgaagttcc tattccgaag ttcctattct 2340
ctagaaagta taggaacttc agagcgctt tgaaaaccaa aagcgctctg aagacgcact 2400
ttcaaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460
tccacaaaca ttgctcaaaa gatatctttt gctatataatc tctgtgctat atcccttat 2520
aacctaccca tccacccccc gtccttgaa cttgcataa aactcgacct ctacatcaac 2580
aggctccaa tgctctcaa attttactgt caagtagacc catacggtctg taatatgctg 2640
ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700
ctttacggtt ccctgagatt gaatttagttt cttagtata tgataacaaga cactttgaa 2760
ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaattt tcctgtcggc 2820
tattatctcc gcctcagttt gatcttccgc ttcagactgc cattttcac ataatgaatc 2880
tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940
atgttgtaca ttgttttagtt cacgagaagg gtcctttca ggcggtagct cctgatctcc 3000
tatatgaccc ttatcctgtt ctcttcacaa aaacttagaa atgtattcat gaattatgga 3060
gcacctaata acattctca aggccggagaa gtttggcca gatgccaaat atgcttgaca 3120
tgaaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180
tcgaagataa gagaagaatg cagtgaccc ttgtattgaca aatggagatt ccatgtatct 3240
aaaaaaatacg cctttaggcc ttctgatacc ctccccctg cggtttagcg tgcctttac 3300
attaatatct aaaccctctc cgatggggc cttaactga ctaataaatacg caaccgatata 3360
aaactgtgat aattctgggt gatttatgat tcgatcgaca attgtattgt acactagtgc 3420
aggatcaggc caatccagtt ctcccccaat taccgggtgt tcgtctgtat tcagtacatg 3480
tccaacaaat gcaaattgcta acgtttgtta tttcttataa ttgtcaggaa ctggaaaagt 3540
cccccttgcc gtctcgatca cacacctact ttcatcgatc accataggtt ggaagtgctg 3600
cataatacat tgcttaatac aagcaagcag tctctgcaca ttcatatttc agttatttc 3660

cattacagct gatgtcattt tatatacgcg ctgtaaaaat ctatctgtta cagaaggttt 3720
tcgcggtttt tataaacaaa actttcggtt caaatcgag caatcacccc agctgcgtat 3780
ttggaaattt gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcatga 3840
ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaac ctcaatctgt 3900
ccattgaatg ccttataaaaa cagctataga ttgcataaaaa gagtttagcta ctcaatgctt 3960
tttgtcaaag cttaactgtatg atgatgtgtc tactttcagg cgggtctgtt gtaaggagaa 4020
tgacattata aagctggcac tttagaattcc acggactata gactatacta gtataactccg 4080
tctactgtac gatacaacttc cgctcaggc tcgttcctt aacgaggcct taccactctt 4140
ttgttactct attgatccag ctcagcaaag gcagtgtat ctaagattct atcttcgcga 4200
tgttagaaaa ctagctagac cgagaaaagag actagaaaatg caaaaaggcac ttctacaatg 4260
gctgccatca ttattatccg atgtgacgt gcattttttt tttttttttt tttttttttt 4320
tttttttttt tttttttttt ttttttgta caaatatcat aaaaaaaagag aatctttta 4380
agcaaggatt ttcttaactt cttcggcgac agcatcacccg acttcgggtgg tactgttgg 4440
accacctaata tcaccagttc tgataacctgc atccaaaacc tttttaactg catcttcaat 4500
ggcttaccc tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560
agtggcgata ggggtgaccc tattcttgg caaatctggc gcggaaaccat ggcattgg 4620
gtacaaaacca aatgcgggtgt tcttgcgtgg caaagaggcc aaggacgcag atggcaacaa 4680
acccaaggag cctgggataa cggaggcttc atcggagatg atatcaccaa acatgttgc 4740
ggtgattata ataccattt agtgggttgg gttcttaactt aggtcatgg cggcagaatc 4800
aatcaatttga tggtaactt tcaatgttgg gaattcggtt ttgatggttt cctccacagt 4860
ttttctccat aatcttgaag aggccaaaac attagcttta tccaaggacc aaataggcaa 4920
tgggtggctca tggtaacttggg ccatgaaagc ggccattttt gtgattctt gcacttctgg 4980
aacgggttat tggtaactt cccaaagcgac accatcacca tggcttccct ttcttccacc 5040
aaagtaaata cctccccacta attctctaac aacaacgaag tcaatgttgg tagcaaattt 5100
tggcttggatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5160
ggcgtacaat tggtaacttggg tacggatttt tagtaaacct tggatggc taaactacc 5220
ggtaaaaaat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttggaggc 5280
ttccagcgcc tcatctggaa gtggacacc tggatggc atagcagcac caccaattaa 5340
atgattttcg aaatcgaaact tggatggc acgaacatca gaaatagcct tggatggc 5400
aatggcttcg gctgtgattt cttgaccaac gtggtcaccc ggcaaaaacga cgatcttctt 5460

cctggccacg gcagaagcac gcttatcgct ccaattccc acaacattag tcaactccgt 7320
taggcccttc attgaaagaa atgaggatcat caaatgtctt ccaatgtgag atttgggcc 7380
atttttata gcaaagattg aataaggcgc attttcttc aaagctttat tgtacgatct 7440
gactaagtta tctttataa attggatttc ctgttattt cttgaagaat tgccggcct 7500
atttactcg ttttaggactg gttcagaatt cctcaaaaat tcataccaaat atacaagtgg 7560
atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7620
atttttata ggtaatgtc atgataataa tggttctta gacgtcagggt ggcactttc 7680
ggggaaatgt gcgcggaacc cctattgtt tattttcta aatacattca aatatgtatc 7740
cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 7800
gtattcaaca tttccgtgtc gcccatttattc cctttttgc ggcattttgc cttccgttt 7860
ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttt ggtgcacgag 7920
tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagttt cgccccgaag 7980
aacgtttcc aatgatgagc acttttaaag ttctgctatg tggcgcggtt ttatcccgt 8040
ttgacgcccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggtt 8100
agtactcacc agtcacagaa aagcatctt cggatggcat gacagtaaga gaattatgca 8160
gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcgag 8220
gaccgaagga gctaaccgct ttttgacaca acatggggta tcatgttaact cgccttgatc 8280
gttggaaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340
cagcaatggc aacaacgttg cgccaaactat taactggcgactacttact ctatccccc 8400
ggcaacaatt aatagactgg atggaggcgg ataaaagttgc aggaccactt ctgcgctcgg 8460
cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggctcgcg 8520
gtatcattgc agcaactgggg ccagatggta agccctcccg tatcgttagtt atctacacga 8580
cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640
tgattaagca ttggtaactg tcagaccaag tttactcata tatacttttag attgatttaa 8700
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgcataat ctcatgacca 8760
aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820
gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaaca aaaaaaccac 8880
cgctaccagc ggtggtttgtt ttgccggatc aagagctacc aactctttt ccgaaggtaa 8940
ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgttagccg tagttaggcc 9000
accacttcaa gaactctgta gcaccgccta cataacctgc tctgctaata ctgttaccag 9060

tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9120
cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180
gaacgaccta caccgaactg agataacctac agcgtgagct atgagaaaagc gccacgcttc 9240
ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaaca ggagagcgca 9300
cgagggagct tccaggggaa aacgccttgtt atctttatag tcctgtcggg tttcggccacc 9360
tctgacttga gcgtcgattt ttgtgtatgtc cgtcaggggg gcggagccta tggaaaaacg 9420
ccagcaacgc ggcctttta cggttcctgg cctttgctg gcctttgct cacatgttct 9480
ttcctgcgtt atcccctgat tctgtggata accgtattac cgccttgag tgagctgata 9540
ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 9600
gcctgatgct gtatttctc cttacgcattc tgtcggtat ttcacaccgc atatggtgc 9660
ctctcagtac aatctgtct gatgcccat agttaagcca gtatacacctc cgctatcgct 9720
acgtgactgg gtcatggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780
ggcttgcgtc ctccggcat ccgcattacag acaagctgtg accgtctccg ggagctgcat 9840
gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctcggtt aaagctcatc 9900
acgcgtggctg tgaagcgatt cacagatgtc tgcctgttca tccgcgttcca gtcgtttag 9960
tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020
ttcctgtttt gtcactgtatg cctccgtgtt agggggattt ctgttcatgg gggtaatgt 10080
accgatgaaa cgagagagga tgctcacat acgggttact gatgtgaac atgcccgggtt 10140
actggAACGT tgtgagggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200
caactcagggt caatgccatc gcttcgtttaa tacagatgtt ggtgttccac aggtagcca 10260
gcagcatcct gcgtatgcaga tccggaacat aatggtgcag ggctgtact tccgcgtttc 10320
cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380
tttgcagcag cagtcgcttc acgttcgttc gcgtatcggtt gattcattct gctaaccagt 10440
aaggcaaccc cgccagccta gcccgggttca aacgcacagg agcacgttca tgcgcacccg 10500
tggccaggac ccaacgctgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10560
gatggatatg ttctgccaag ggttggtttt cgcattcaca gttctccgca agaatttattt 10620
ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcagggtc 10680
gaggtggccc ggctccatgc accgcgacgc aacgcggggaa ggcagacaag gtatagggcg 10740
gcgcctacaa tccatgccaat cccgttccat gtgctcgccg aggccggata aatcgccgtg 10800
acgtcagcg gtccaatgtat cgaagtttagg ctggtaagag ccgcgagcga tccttgaagc 10860

tgtccctgat ggtcgcatc tacctgcctg gacagcatgg cctgcaacgc gggcatccc 10920
atgccgcgg aagcgagaag aatcataatg gggaggcca tccagcctcg cgtcgcgaac 10980
gccagcaaga cgtagcccg cgcgtcggcc gccatgcggg cgataatggc ctgcttctcg 11040
ccgaaacgtt tggggcggg accagtgacg aaggcttgag cgagggcgtg caagattccg 11100
aataccgcaa gcgacaggcc gatcatgtc ggcgtccagc gaaagcggtc ctcgcccggaaa 11160
atgaccaga gcgcgtccgg cacctgtcct acgagttgca tgataaaagaa gacagtcata 11220
agtgcggcga cgatagtcat gccccggcc caccggagg agctgactgg gttgaaggct 11280
ctcaaggcga tcggtcgagg atccttcaat atgcgcacat acgctgttat gttcaaggctc 11340
ccttcgttta agaacgaaag cggcttcct tttgagggat gttcaagtt gttcaaatct 11400
atcaaatttg caaatcccc agtctgtatct agagcgttga atcggtgatg cgatttgtta 11460
attaaattga tgggtgtcacc attaccaggc ttagatatac caatggcaaa ctgagcacaa 11520
caataccagt ccggatcaac tggcaccatc tctccgtag tctcatctaa tttttcttcc 11580
ggatgaggtt ccagatatac cgcaacacct ttattatggt tccctgagg gaataataga 11640
atgtcccatt cgaaatcacc aattctaaac ctggcgaat tgtatccgg gtttgttaac 11700
tcgttccagt caggaatgtt ccacgtgaag ctatcttcca gcaaagtctc cacttcttca 11760
tcaaattgtg gagaatactc ccaatgtct tatctatggg acttccggg aacacagtac 11820
cgatacttcc caattcgtct tcagagctca ttgtttgtt gaagagacta atcaaagaat 11880
cgaaaaatcaa tatcttaact gatagttga tcaaaggggc aaaacgtagg 11940
ggcaaacaaa cggaaaaatc gtttctaaa ttttctgatg ccaagaactc taaccagtct 12000
tatctaaaaa ttgccttatg atccgtctct ccgggttacag cctgtgttaac tgattaatcc 12060
tgcctttcta atcaccattc taatgtttta attaagggat ttgttcttca ttaacggctt 12120
tcgctcataa aaatgttatg acgtttgcc cgcaggcggg aaaccatcca cttcacgaga 12180
ctgatctcct ctgccggAAC accggggcatc tccaacttat aagttggaga aataagagaa 12240
tttcagattt agagaatgaa aaaaaaaaaac ctttagttca taggtccatt ctcttagcgc 12300
aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtgtatg 12360
caacctgcct ggagtaatg atgacacaag gcaattgacc cacgtatgtc tctatctcat 12420
tttcttacac cttcttattac cttctgtct ctctgatttg gaaaaagctg aaaaaaaaaagg 12480
ttgaaaccag ttccctgaaa ttattccct acttgactaa taagtatata aagacggtag 12540
gtattgattt gtaattctgtt aatctatttc ttaaacttct taaattctac ttttataatgtt 12600
agtctttttt ttagtttaa aacaccaaga acttagttc gaataaacac acataaacaa 12660

acaagcttac aaaacaaa atg gct gca tat gca gct cag ggc tat aag gtg 12711
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val
 1 5 10

cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac 12759
 Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr
 15 20 25

atg tcc aag gct cat ggg atc gat cct aac atc agg acc ggg gtg aga 12807
 Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg
 30 35 40

aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc 12855
 Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe
 45 50 55

ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt 12903
 Leu Ala Asp Gly Gly Cys Ser Gly Ala Tyr Asp Ile Ile Ile Cys
 60 65 70 75

gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc att ggc act 12951
 Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr
 80 85 90

gtc ctt gac caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc gcc 12999
 Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala
 95 100 105

acc gcc acc cct ccg ggc tcc gtc act gtg ccc cat ccc aac atc gag 13047
 Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu
 110 115 120

gag gtt gct ctg tcc acc acc gga gag atc cct ttt tac ggc aag gct 13095
 Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala
 125 130 135

atc ccc ctc gaa gta atc aag ggg ggg aga cat ctc atc ttc tgt cat 13143
 Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His
 140 145 150 155

tca aag aag aag tgc gac gaa ctc gcc gca aag ctg gtc gca ttg ggc 13191
 Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly
 160 165 170

atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg tcc gtc atc ccg 13239
 Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro
 175 180 185

acc agc ggc gat gtt gtc gtc gtg gca acc gat gcc ctc atg acc ggc 13287
 Thr Ser Gly Asp Val Val Val Ala Thr Asp Ala Leu Met Thr Gly
 190 195 200

tat acc ggc gac ttc gac tcg gtg ata gac tgc aat acg tgt gtc acc 13335
 Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr
 205 210 215

cag aca gtc gat ttc agc ctt gac cct acc ttc acc att gag aca atc 13383
 Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile
 220 225 230 235

acg ctc ccc caa gat gct gtc tcc cgc act caa cgt cgg ggc agg act	13431
Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr	
240	245
250	
ggc agg ggg aag cca ggc atc tac aga ttt gtg gca cgg ggg gag cgc	13479
Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg	
255	260
265	
ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gca	13527
Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala	
270	275
280	
ggc tgt gct tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta	13575
Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu	
285	290
295	
cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat ctt	13623
Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu	
300	305
310	315
gaa ttt tgg gag ggc gtc ttt aca ggc ctc act cat ata gat gcc cac	13671
Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His	
320	325
330	
ttt cta tcc cag aca aag cag agt ggg gag aac ctt cct tac ctg gta	13719
Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val	
335	340
345	
gcg tac caa gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg	13767
Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser	
350	355
360	
tgg gac cag atg tgg aag tgt ttg att cgc ctc aag ccc acc ctc cat	13815
Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His	
365	370
375	
ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa atc	13863
Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile	
380	385
390	395
acc ctg acg cac cca gtc acc aaa tac atc atg aca tgc atg tcg gcc	13911
Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala	
400	405
410	
gac ctg gag gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg	13959
Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu	
415	420
425	
gct gct ttg gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata gtg	14007
Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val	
430	435
440	
ggc agg gtc gtc ttg tcc ggg aag ccg gca atc ata cct gac agg gaa	14055
Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu	
445	450
455	
gtc ctc tac cga gag ttc gat gag atg gaa gag tgc tct cag cac tta	14103
Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu	
460	465
470	475

ccg tac atc gag caa ggg atg atg ctc gcc gag cag ttc aag cag aag 14151
 Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys
 480 485 490

 gcc ctc ggc ctc ctg cag acc gcg tcc cgt cag gca gag gtt atc gcc 14199
 Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala
 495 500 505

 cct gct gtc cag acc aac tgg caa aaa ctc gag acc ttc tgg gcg aag 14247
 Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys
 510 515 520

 cat atg tgg aac ttc atc agt ggg ata caa tac ttg gcg ggc ttg tca 14295
 His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser
 525 530 535

 acg ctg cct ggt aac ccc gcc att gct tca ttg atg gct ttt aca gct 14343
 Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala
 540 545 550 555

 gct gtc acc agc cca cta acc act agc caa acc ctc ctc ttc aac ata 14391
 Ala Val Thr Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile
 560 565 570

 ttg ggg ggg tgg gtg gct gcc cag ctc gcc ccc ggt gcc gct act 14439
 Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr
 575 580 585

 gcc ttt gtg ggc gct ggc tta gct ggc gcc atc ggc agt gtt gga 14487
 Ala Phe Val Gly Ala Gly Leu Ala Ala Ile Gly Ser Val Gly
 590 595 600

 ctg ggg aag gtc ctc ata gac atc ctt gca ggg tat ggc gcg ggc gtg 14535
 Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val
 605 610 615

 gcg gga gct ctt gtg gca ttc aag atc atg agc ggt gag gtc ccc tcc 14583
 Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser
 620 625 630 635

 acg gag gac ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc gga gcc 14631
 Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala
 640 645 650

 ctc gta gtc ggc gtg gtc tgt gca gca ata ctg cgc cg^g cac gtt ggc 14679
 Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly
 655 660 665

 ccg ggc gag ggg gca gtg cag tgg atg aac cgg ctg ata gcc ttc gcc 14727
 Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala
 670 675 680

 tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat 14775
 Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp
 685 690 695

 gca gct gcc cgc gtc act gcc ata ctc agc agc ctc act gta acc cag 14823
 Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln
 700 705 710 715

ctc ctg agg cga ctg cac cag tgg ata agc tcg gag tgt acc act cca Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro 720 725 730	14871
tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val 735 740 745	14919
ttg agc gac ttt aag acc tgg cta aaa gct aag ctc atg cca cag ctg Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu 750 755 760	14967
cct ggg atc ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp 765 770 775	15015
cga ggg gac ggc atc atg cac act cgc tgc cac tgt gga gct gag atc Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile 780 785 790 795	15063
act gga cat gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr 800 805 810	15111
tgc agg aac atg tgg agt ggg acc ttc ccc att aat gcc tac acc acg Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr 815 820 825	15159
ggc ccc tgt acc ccc ctt cct gcg ccg aac tac acg ttc gcg cta tgg Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp 830 835 840	15207
agg gtg tct gca gag gaa tac gtg gag ata agg cag gtg ggg gac ttc Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe 845 850 855	15255
cac tac gtg acg ggt atg act act gac aat ctt aaa tgc ccg tgc cag His Tyr Val Thr Gly Met Thr Asp Asn Leu Lys Cys Pro Cys Gln 860 865 870 875	15303
gtc cca tcg ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cat Val Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His 880 885 890	15351
agg ttt gcg ccc ccc tgc aag ccc ttg ctg cgg gag gag gta tca ttc Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe 895 900 905	15399
aga gta gga ctc cac gaa tac ccg gta ggg tcg caa tta cct tgc gag Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu 910 915 920	15447
ccc gaa ccg gac gtg gcc gtg ttg acg tcc atg ctc act gat ccc tcc Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser 925 930 935	15495
cat ata aca gca gag ggc ggg cga agg ttg gcg agg gga tca ccc His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro 940 945 950 955	15543

ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc Pro Ser Val Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu 960 965 970	15591
aag gca act tgc acc gct aac cat gac tcc cct gat gct gag ctc ata Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile 975 980 985	15639
gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg 990 995 1000	15687
gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu 1005 1010 1015	15735
gtg gcg gag gag gac gag cg ^g gag atc tcc gta ccc gca gaa atc ctg Val Ala Glu Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu 1020 1025 1030 1035	15783
cgg aag tct cgg aga ttc gcc cag gcc ctg ccc gtt tgg gcg cg ^g cg ^g Arg Lys Ser Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro 1040 1045 1050	15831
gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu 1055 1060 1065	15879
cca cct gtg gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro 1070 1075 1080	15927
gtg cct ccg cct ccg aag aag ccg acg gtg gtc ctc act gaa tca acc Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr 1085 1090 1095	15975
cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser 1100 1105 1110 1115	16023
tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro 1120 1125 1130	16071
gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser 1135 1140 1145	16119
atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly 1150 1155 1160	16167
tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys 1165 1170 1175	16215
tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala 1180 1185 1190 1195	16263

gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu 1200 1205 1210	16311
cgt cac cac aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln 1215 1220 1225	16359
agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His 1230 1235 1240	16407
tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ser Lys Val Lys 1245 1250 1255	16455
gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His 1260 1265 1270 1275	16503
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His 1280 1285 1290	16551
gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu 1295 1300 1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1310 1315 1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1325 1330 1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1340 1345 1350 1355	16743
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr 1360 1365 1370	16791
gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala 1375 1380 1385	16839
tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1390 1395 1400	16887
ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca atc Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile 1405 1410 1415	16935
tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser 1420 1425 1430 1435	16983

ctc acc gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg 17031
 Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly
 1440 1445 1450

gag aac tgc ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act 17079
 Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr
 1455 1460 1465

agc tgt ggt aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt 17127
 Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys
 1470 1475 1480

cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac 17175
 Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp
 1485 1490 1495

tta gtc gtt atc tgt gaa agc gcg ggg gtc cag gag gac gcg gcg agc 17223
 Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser
 1500 1505 1510 1515

ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg 17271
 Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly
 1520 1525 1530

gac ccc cca caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc 17319
 Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser
 1535 1540 1545

tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac 17367
 Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr
 1550 1555 1560

ctc acc cgt gac cct aca acc ccc ctc gcg aga gct gcg tgg gag aca 17415
 Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr
 1565 1570 1575

gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt 17463
 Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe
 1580 1585 1590 1595

gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc 17511
 Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser
 1600 1605 1610

gtc ctt ata gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc 17559
 Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile
 1615 1620 1625

tac ggg gcc tgc tac tcc ata gaa cca ctg gat cta cct cca atc att 17607
 Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile
 1630 1635 1640

caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca 17655
 Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro
 1645 1650 1655

ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg 17703
 Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro
 1660 1665 1670 1675

ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt 17751
 Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu
 1680 1685 1690

 ctg gcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac 17799
 Leu Ala Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn
 1695 1700 1705

 tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc 17847
 Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly
 1710 1715 1720

 cag ctg gac ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac 17895
 Gln Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp
 1725 1730 1735

 att tat cac agc gtg tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc 17943
 Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys
 1740 1745 1750 1755

 cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga 17991
 Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1760 1765 1770

 atg agc acg aat cct aaa cct caa aga aag acc aaa cgt aac acc aac 18039
 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
 1775 1780 1785

 cgg cgx ccg cag gac gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt 18087
 Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
 1790 1795 1800

 gga gtt tac ttg ttg ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg 18135
 Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
 1805 1810 1815

 acg aga aag act tcc gag ccg tcg caa cct cga ggt aga cgt cag cct 18183
 Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro
 1820 1825 1830 1835

 atc ccc aag gct cgt ccg ccc gag ggc agg acc tgg gct cag ccc ggg 18231
 Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly
 1840 1845 1850

 tac cct tgg ccc ctc tat ggc aat gag ggc tgc ggg tgg gcg gga tgg 18279
 Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp
 1855 1860 1865

 ctc ctg tct ccc cgt ggc tct cgg cct agc tgg ggc ccc aca gac ccc 18327
 Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
 1870 1875 1880

 cgx cgt agg tcg cgc aat ttg ggt aag gtc atc gat acc ctt acg tgc 18375
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
 1885 1890 1895

 ggc ttc gcc gac ctc atg ggg tac ata ccg ctc gtc ggc gcc cct ctt 18423
 Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu
 1900 1905 1910 1915

gga ggc gct gcc agg gcc ctg gcg cat ggc gtc cgg gtt ctg gaa gac 18471
Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp
1920 1925 1930
ggc gtg aac tat gca aca ggg aac ctt cct ggt tgc tct taatagtcga 18520
Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser
1935 1940
ctttgttccc actgtacttt tagctcgtaaaaatacaat atactttca tttctccgtat 18580
aacaacatgt ttccccatgt aatatccttt tctattttc gttccgttac caactttaca 18640
catactttat atagctattc acttctatac actaaaaaaac taagacaatt ttaattttgc 18700
tgccctgccat atttcaattt gttataaattt cctataattt atcctatttag tagctaaaaaa 18760
aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggc 18820
gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tggcgccgg 18880
ccaaagcggt cggacagtgc tccgagaacg ggtgcgcata gaaattgcata caacgcataat 18940
agcgctagca gcacgcccata gtgactggcg atgctgtcgaaatggacgat atcccgcaag 19000
aggccccggca gtaccggcat aaccaaggct atgcctacag catccagggt gacggtgccg 19060
aggatgacga tgagcgcattt gtttagatttccatcacggtg cctgactgcgttagcaattt 19120
aactgtgata aactaccgca ttaaagctttt ttctttccaa tttttttttt ttcgtcatta 19180
taaaaaatcat tacgaccgag attcccggtt aataactgat ataattaaat tgaagctcta 19240
atttgtgagt ttagtataca tgcatttact tataatacag ttttttagttt ttgctggccg 19300
catcttcctca aatatgcttc ccagcctgctt tttctgtac gttcacccctc taccttagca 19360
tcccttcctt ttgcaaatacg tccctttccaa acaataataa tgcacatcc tgcacatcc 19420
acatcatcca cgggtctata ctgttgaccc aatgcgtctc cttgtcatc taaacccaca 19480
ccgggtgtca taatcaacca atcgtaacct tcatacttc caccatgtc tctttgagca 19540
ataaaagccga taacaaaatc tttgtcgctc ttgcataatgt caacagtacc ctttagtat 19600
tctccagtag atagggagcc cttgcatttgc aatttgcata acatcaaaag gcctcttaggt 19660
tccctttgtta cttcttcgtc cgcctgcttc aaaccgctaa caatacctgg gcccaccaca 19720
ccgtgtgtcat tcgtaatgtc tgccctttctt gctattctgtt atacacccgc agagtagtgc 19780
aatttgcatttgc tattaccaat gtcagcaaat tttctgtctt cgaagagtaa aaaattgtac 19840
ttggcggata atgccttttag cggcttaact gtgcctccaa tggaaaaatc agtcaagata 19900
tccacatgttgc ttttttagtaa acaaatttttggacctaacttgc cttcaactaa cttccagtaat 19960
tcccttgggtgg tacgaacatc caatgaagca cacaagtttgc tttgtttttc gtgcattgtata 20020
ttaaaatagct tggcagcaac aggacttagga tgagtagcag cacgttcctt atatgtat 20080

ttcgacatga tttatcttcg tttcctgcag gttttgttc tgtgcagttg ggtaagaat 20140
 actggcaat ttcatgttcc ttcaacacta catatgcgta tatataccaa tctaagtctg 20200
 tgctccttcc ttcggttcc cttctgttcc gagattaccg aatcaaaaaa atttcaagga 20260
 aaccgaaatc aaaaaaaaaga ataaaaaaaaa aatgtgaat tgaaaagctt atcgat 20316

<210> 15
 <211> 1944
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core173

<400> 15
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
 1 5 10 15
 Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
 20 25 30
 Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
 35 40 45
 Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
 50 55 60
 Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser
 65 70 75 80
 Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
 85 90 95
 Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
 100 105 110
 Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
 115 120 125
 Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
 130 135 140
 Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys
 145 150 155 160
 Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala
 165 170 175
 Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
 180 185 190
 Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe
 195 200 205
 Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe

210	215	220
Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp		
225	230	235
Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro		
245	250	255
Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe		
260	265	270
Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr		
275	280	285
Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn		
290	295	300
Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly		
305	310	315
Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr		
325	330	335
Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr		
340	345	350
Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp		
355	360	365
Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu		
370	375	380
Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro		
385	390	395
Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val		
405	410	415
Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala		
420	425	430
Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu		
435	440	445
Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu		
450	455	460
Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln		
465	470	475
Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu		
485	490	495
Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr		
500	505	510
Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe		
515	520	525

Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
 530 535 540

Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
 545 550 555 560

Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575

Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590

Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605

Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
 610 615 620

Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
 625 630 635 640

Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655

Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
 660 665 670

Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
 675 680 685

Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
 690 695 700

Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
 705 710 715 720

His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735

Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750

Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765

Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
 770 775 780

Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
 785 790 795 800

Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
 805 810 815

Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
 820 825 830

Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
 835 840 845

Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
 850 855 860

 Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
 865 870 875 880

 Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
 885 890 895

 Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910

 Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
 915 920 925

 Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
 930 935 940

 Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
 945 950 955 960

 Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975

 Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
 980 985 990

 Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
 995 1000 1005

 Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
 1010 1015 1020

 Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
 1025 1030 1035 1040

 Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
 1045 1050 1055

 Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070

 Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
 1075 1080 1085

 Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
 1090 1095 1100

 Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
 1095 1110 1115 1120

 Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135

 Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
 1140 1145 1150

 Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
 1155 1160 1165

Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
 1170 1175 1180

Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 185 1190 1195 1200

Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
 1205 1210 1215

Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
 1220 1225 1230

Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
 1235 1240 1245

Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
 1250 1255 1260

Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 265 1270 1275 1280

Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
 1285 1290 1295

Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310

Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325

Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 1330 1335 1340

Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
 345 1350 1355 1360

Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
 1365 1370 1375

Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
 1380 1385 1390

Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
 1395 1400 1405

Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
 1410 1415 1420

Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
 425 1430 1435 1440

Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
 1445 1450 1455

Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470

Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
 1475 1480 1485

Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
 1490 1495 1500

Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
 505 1510 1515 1520

Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535

Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
 1540 1545 1550

Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
 1555 1560 1565

Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
 1570 1575 1580

Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
 585 1590 1595 1600

Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
 1605 1610 1615

Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
 1620 1625 1630

Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
 1635 1640 1645

Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
 1650 1655 1660

Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
 665 1670 1675 1680

Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
 1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser
 1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
 1730 1735 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala
 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro
 1765 1770 1775

Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
 1780 1785 1790

Val Lys Phe Pro Gly Gly Gln Ile Val Gly Val Tyr Leu Leu
 1795 1800 1805

Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
 1810 1815 1820

Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
 825 1830 1835 1840

Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
 1845 1850 1855

Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
 1860 1865 1870

Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
 1875 1880 1885

Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
 1890 1895 1900

Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
 905 1910 1915 1920

Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
 1925 1930 1935

Thr Gly Asn Leu Pro Gly Cys Ser
 1940

<210> 16

<211> 20217

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core140

<220>

<221> CDS

<222> (12679)..(18411)

<400> 16

atcgatccta ccccttgcgc taaaagaagta tatgtgccta ctaacgcttg tctttgtctc 60

tgtcaactaaa cactggatta ttactcccag atacttattt tggactaatt taaatgattt 120

cggatcaacg ttcttaatat cgctgaatct tccacaattt atgaaaagtag ctaggaagag 180

gaattggtat aaagtttttg tttttgtaaa tctcgaagta tactcaaacg aatttagtat 240

tttctcagtg atctcccaga tgctttcacc ctcacttaga agtgctttaa gcatttttt 300

actgtggcta tttcccttat ctgcttcctc cgatgattcg aactgtaatt gcaaactact 360

tacaatatca gtgatatacg attgatgttt ttgtccatag taaggaataa ttgtaaattc 420

ccaaggcagga atcaatttct ttaatgaggc ttccagaatt gttgctttt gcgtcttgta 480

tttaaactgg agtgatattat tgacaatatac gaaactcagc gaattgctta tgatagtatt 540

atagctcatg aatgtggctc tcttgattgc tgccccgtta tgtgtaatca tccaaacataa 600
ataggtagt tcagcagcac ataatgctat ttttcacact gaaggcttt caaaccttc 660
cacaaaactga cgaacaagca ccttaggtgg tggtttacat aatatataa attgtggcat 720
gcttagcgcc gatcttgtgt gcaattgata tctagttca actactctat ttatcttgc 780
tcttgagta ttcaaacadg ctaactcgaa aaactaactt taattgtcct gttgtctcg 840
cgttcttcg aaaaatgcac cggccgcgca ttattgtac tgcggaaaata attggactg 900
cggtatcttc attcatatt taaaaatgc accttgcgtg ctttccctt attttagac 960
ggcccccgagg ttcgtttgc ggtactatct tgtgataaaa agttgtttt acatgtgatc 1020
tgcacagatt ttataatgta ataagcaaga atacatttac aaacgaacaa tactggtaaa 1080
agaaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140
cttatagcgt ctggatgta tgcggctgt ttattgaaat gattgctcct gatgtagata 1200
ttgatataga gagtaaacgt aagtctgatg agctactctt tccaggatat gtcataaggc 1260
ccatggaaatc tctcacaacc ggtaggccgt atggcttga ttctagcgca gaagattcca 1320
gcgtatcttc tgactccagt gctgaggtaa tttgcctgc tgcgaagatg gttaaggaaa 1380
ggtttgattc gattggaaat ggtatgctct cttcacaaga agcaagtca gctgccatag 1440
atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500
ctataataat aggaagattg cccgagaaag acaagaagag agctaccgaa atgctcatga 1560
gaaaaatgga ttgtacacag ttattagtc caccagctcc aacggaagaa gatgttatga 1620
agctcgtaag cgtcgttacc caattgcctt cttagttcc accagatcgt caagctgctt 1680
taataggtga tttattcattc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740
tggccgcaga gaatcgaaa cagaaaaaaaaa agagttagt ggaaggaagg actgaagtga 1800
accatgctaa tacaatgaa gaagttccct ccaggcgaac aagaagttaga gacacaaatg 1860
caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacga 1920
aaaaaaaggag agtagcaacg agggtaaggg gcagaaaaatc acgtaataact tctagggat 1980
gatccaatat caaaggaaat gatgcattt aaggatgaga ctaatccat tgaggagtgg 2040
cagcatatag aacagctaaa gggtagtgcgta agggtaagca tacgatacccg catggat 2100
gggataatat cacaggaggt actagactac cttagtccat acataaatacg acgcatataa 2160
gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220
caacacgcag atataggtgc gacgtgaaca gtgagctgta tgcgcgcagc tcgcgttgca 2280
tttcggaaag cgctcgaaaa cgaaaacgct ttgaagttcc tattccgaag ttcctattct 2340

ctagaaaagta taggaacttc agagcgctt tgaaaaccaa aagcgctctg aagacgcact 2400
ttcaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460
tccacaaaca ttgctcaaaa gtatctctt gctatatatc tctgtgctat atccctatat 2520
aacctaccca tccacaccttc gtccttgaa cttgcatact aactcgacct ctacatcaac 2580
aggcttccaa tgctttcaa attttactgt caagtagacc catacggtcg taatatgctg 2640
ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700
ctttacggtt ccctgagatt gaatttagttc cttagtata tgataacaaga cactttgaa 2760
ctttgtacga cgaattttga ggttcgccc catctggcta tttccaatta tcctgtcgcc 2820
tattatctcc gcctcagttt gatcttccgc ttcagactgc cattttcac ataatgaatc 2880
tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940
atgttgtaca ttgttttagtt cacgagaagg gtcctttca ggcggtagct cctgatctcc 3000
tatatgacct ttatcctgtt ctcttccac aaacttagaa atgtattcat gaattatgga 3060
gcacctaata acattcttca aggccggagaa gtttggcca gatgcccaat atgcttgaca 3120
tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180
tcgaagataa gagaagaatg cagtgcaccc ttgtattgaca aatggagatt ccatgtatct 3240
aaaaaaatacg cctttaggcc ttctgatacc ctccccctg cggttagcg tgcctttac 3300
attaatatct aaaccctctc cgatggggc cttaactga ctaataaaatg caaccgatat 3360
aaactgtgat aattctgggt gatttatgat tcgatcgaca attgtattgt acactagtgc 3420
aggatcagggc caatccagtt cttttcaat taccgggtgt tcgtctgtat tcagtacatg 3480
tccaacaaat gcaaattgcta acgtttgtt tttcttataa ttgtcaggaa ctggaaaagt 3540
cccccttgc gtctcgatta cacacctact ttcatcgta accataggtt ggaagtgcgtg 3600
cataatacat tgcttaatac aagcaagcag tctctcgcca ttcatatttc agttatttc 3660
cattacagct gatgtcattt gatgtcattt ctgtaaaaat ctatctgtt cagaaggttt 3720
tcgcgggttt tataaacaaa actttcgta cgaaatcgag caatcacccc agctgcgtat 3780
ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcata 3840
ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaac ctcaatctgt 3900
ccattgaatg ctttataaaa cagctataga ttgcataagaa gagttgcata ctcaatgctt 3960
tttgcgtcaag cttactgtatg atgatgtgtc tactttcagg cgggtctgtt gtaaggagaa 4020
tgacattata aagctggcac tttagaattcc acggactata gactatacta gtataactccg 4080
tctactgtac gatacacttc cgctcaggc cttgtcctt aacgaggcct taccacttt 4140

ttgttactct attgatccag ctcagcaaag gcagtgtat ctaagattct atcttcgcga 4200
tgtagtaaaa ctagctagac cgagaaagag actagaaatg caaaaggcac ttctacaatg 4260
gctgccatca ttattatccg atgtgacgct gcatttttt tttttttttt 4320
ttttttttt tttttttttt ttttttgta caaatatcat aaaaaaaagag aatctttta 4380
agcaaggatt ttcttaactt ctgcggcgac agcatcaccg acttcggtgg tactgttgaa 4440
accacctaata tcaccagttc tgataacctgc atccaaaacc ttttaactg catcttaat 4500
ggctttacct tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560
agtggcgata ggggtgaccc tattctttgg caaatctggc gcggaaccat ggcatggttc 4620
gtacaaacca aatgcggtgt tcttgcgtgg caaagaggcc aaggacgcag atggcaacaa 4680
acccaaggag cctgggataa cgaggcgatc atcggagatg atatcaccaa acatgttgct 4740
ggtgattata ataccatata ggtgggttgg gttcttaact aggtatcatgg cgccagaatc 4800
aatcaattga tggtaactt tcaatgttagg gaattcggtc ttgatggttt cttccacagt 4860
ttttctccat aatcttgaag aggccaaaac attagctta tccaggacc aaataggcaa 4920
tggggctca tggtaaggccatgaaagc ggccattctt gtgattctt gcacttctgg 4980
aacgggttat tggtaactat cccaaagcgac accatcacca tcgtcttcct ttcttttacc 5040
aaagtaataa cttccacta attctctaacc aacaacgaag tcagttaccc tagcaaattg 5100
tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5160
ggcgtacaat tgaagttctt tacggatttt tagtaaacct tggtaggtc taacactacc 5220
ggtaaaaaat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttgaggc 5280
ttccagcgcc tcatactggaa gtggAACACC tggtagcatcg atagcagcac caccaattaa 5340
atgattttcg aaatcgaact tgacatttggaa acgaacatca gaaatagctt taagaacctt 5400
aatggcttcg gctgtgattt cttgaccaac gtggcacctt ggcaaaacga cgatcttctt 5460
aggggcagac attacaatgg tatatccttg aaatataat aaaaaaaaaa aaaaaaaaaa 5520
aaaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5580
tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640
acatccgaac ctggagttt tccctgaaac agatagtata tttgaacctg tataataata 5700
tatagtcgtt cgttttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5760
atctattgca taggtatatct tgcacgtcgc atccccgggtt cattttctgc gtttccatct 5820
tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaataa 5880
atgcaacgcgc agagcgctaa ttttcaaac aaagaatctg agctgcattt ttacagaaca 5940

gaaatgcAAC gcgaaAGCgc tattttacca acgaagaatc tgtgcttcAT ttttGtaaaa 6000
caaaaaatgca acgcgagAGC gctaatttt caaacAAAGA atctgagCTG cattttaca 6060
gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttctttttg 6120
ttctacaaaa atgcatcccG agagcgctat ttttctaaca aagcatctta gattacttt 6180
tttctccTTT gtgcgctcta taatgcagTC tcttgataac ttttgcact gtaggtccgt 6240
taaggtaga agaaggctac tttgggtgtct attttctt ccataaaaaaa agcctgactc 6300
cacttcccgc gtttactgat tactagcgaa gctgcgggtg catttttca agataaaggc 6360
atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6420
gcgttcatGA ttcttcattt gtcagaaaa tatgaacggt ttcttctatt ttgtctctat 6480
atactacgta taggaaatgt ttacatttc gtattgtttt cgattcactc tatgaatagt 6540
tcttactaca attttttGT ctAAAGAGTA atactagaga taaacataaa aaatgttagAG 6600
gtcgagtttA gatgcAAgTT caaggAGCGA aaggTggATg ggtaggTTat atagggatAT 6660
agcacagaga tatatAGcaa agagataACTT ttgagcaATg tttgtggAAg cggTattcgc 6720
aatattttAG tagtcgttA cagtccggTG cgTTTTGGT ttttGAAAG tgcgtcttca 6780
gagcgctttt ggTTTcaAA agcgctctGA agttcctata ctTTctAGAG aatAGGAact 6840
tcggaatAGG aacttcaaAG cgTTTCCGAA aacgagcgcT tccgAAAATg caacgcgAGC 6900
tgcgcacata cagtcactg ttacgcgc ACCTATATCT gcgtgttGCC tGTatATA 6960
tatacatgag aagaACGGCA tagtgcgtgt ttatgcttaA atgcgtactt atatgcgtct 7020
atTTatgtAG gatgaaAGGT agtctAGTAC ctccTGTGAT attatccCAT tccatgcggg 7080
gtatcgtatG ctTccTTcAg cactaccTT tagctgtct atatgctGCC actcctcaat 7140
tggattAGTC tcAtcCttCA atgcTatCAT ttccTTGAT attggatCAT atgcataGTA 7200
ccgagaaaACT agtgcgaAGT agtgcgtAGG tattgcgtt atctgatGAG tatacgttgt 7260
cctggccacG gcagaAGCac gcttatcgct ccaatttccc acaacattAG tcaactccgt 7320
taggcccTTc attgaaAGAA atgaggTCat caaatgtctt ccaatgtgag atTTgggCC 7380
atTTTTATA gcaaAGATTG aataAGGCgc atTTTCTTC aaagctttat tgtacgatct 7440
gactaAGTtA tctttAATA attggTattc ctgtttattG cttgaagaat tgccggcct 7500
atTTactcgt tttaggactG gttcagaatt cctcaaaaAT tcAtccAAAT atacaAGTgg 7560
atcgatgata agctgtcaAA catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7620
tatTTTTATA ggttaatgtc atgataataA tggTTTCTTA gacgtcaggT ggcactttc 7680
ggggAAATgt gcgcggAAacc cctatttGTT tattttctA aatacattca aatATgttAc 7740

cgctcatgag acaataaccc tgataaaatgc ttcaataata ttgaaaaagg aagagtatga 7800
gtattcaaca tttccgtgtc gcccatttc cctttttgc ggcattttgc cttcctgttt 7860
ttgctcaccc agaaacgctg gtgaaaagtaa aagatgctga agatcagttt ggtgcacgag 7920
tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagttt cgccccgaag 7980
aacgtttcc aatgatgagc acttttaaag ttctgctatg tggcgcggtt ttatcccgtg 8040
ttgacgcccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggtt 8100
agtactcacc agtcacagaa aagcatcttta cgatggcat gacagtaaga gaattatgca 8160
gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcgag 8220
gaccgaagga gctaaccgct tttttgcaca acatggggga tcataact cgccttgatc 8280
gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340
cagcaatggc aacaacgttg cgcaaaactat taactggcga actacttact cttagttccc 8400
ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 8460
cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggctcgcg 8520
gtatcattgc agcaactgggg ccagatggta agccctcccg tatcgttagt atctacacga 8580
cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640
tgattaagca ttggtaactg tcagaccaag tttactcata tatacttttag attgatttaa 8700
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttataat ctcatgacca 8760
aaatccctta acgtgagttt tcgttccact gagcgtcaga cccctgtagaa aagatcaaag 8820
gatcttcttgc agatcctttt tttctgcgcg taatctgctg cttgcaaaaca aaaaaaccac 8880
cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactctttt ccgaaggtaa 8940 .
ctggcttcag cagagcgcag ataccaaata ctgtccttct agttagccg tagttaggcc 9000
accacttcaa gaactctgta gcaccgccta cataacctgc tctgctaattc ctgttaccag 9060
tggctgctgc cagtgccgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9120
cggataaggc gcagcggcgcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180
gaacgaccta caccgaactg agataacctac agcgtgagct atgagaaagc gccacgcctc 9240
ccgaaggggag aaaggcggac aggtatccgg taagcggcag ggtcgaaaca ggagagcgca 9300
cgagggagct tccagggggaa aacgccttgtt atctttatag tcctgtcggg tttcgccacc 9360
tctgacttga gcgtcgattt ttgtgatgtc cgtcaggggg gcggagccta tggaaaaacg 9420
ccagcaacgc ggcctttta cggttcctgg cctttgctg gcctttgct cacatgttct 9480
ttcctgcgtt atccccctgat tctgtggata accgtattac cgcctttgag tgagctgata 9540

ccgctcgccc cagccgaacg accgagcgcgca gcgagtcagt gagcgaggaa gcggaaagagc 9600
gcctgatgcg gtattttctc cttacgcac tgcgttat ttcacaccgc atatggtgc 9660
ctctcagtac aatctgctc gatgccgcat agttaagcca gtatacactc cgctatcgct 9720
acgtgactgg gtcatggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780
ggcttgctg ctccggcat ccgccttacag acaagctgtg accgtctccg ggagctgcat 9840
tgtcagagg tttcaccgt catcaccgaa acgcgcgagg cagctgcgtt aaagctcatc 9900
agcgtggctg tgaagcgatt cacagatgtc tgcctgttca tccgcgtcca gctcgtttag 9960
tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020
ttcctgtttg gtcaactgatg cctccgtgt aaaaaaaaaaaaaaaaat 10080
accgatgaaa cgagagagga tgctcacat acgggttact gatgatgaac atgcccgtt 10140
actggaaacgt tgtgagggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200
caactcagggt caatgccagc gcttcgttac tacagatgtc ggtgttccac aggtagcc 10260
gcagcatcct gcgtgcaga tccggAACAT aatggtgcatgg ggcgcgtact tccgcgtttc 10320
cagactttac gaaacacgg aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380
tttgccaggcagc cagtcgcttc acgttcgttc gctatcggtt gattcattct gctaaccagt 10440
aaggcaaccc cggccggcttca aacgcacagg agcacaatca tgcgcacccg 10500
tggccaggac ccaacgcgtgc cccggatgcg cccgcgtgcgg ctgctggaga tggcggacgc 10560
gatggatatg ttctgccaag gttggtttgcgatttca gttctccgc 10620
ggctccaattt cttggagtgg tgaatccgtt agcgagggtgc cccggcttc cattcagggtc 10680
gaggtggccc ggctccatgc accgcgcacgc aacgcgggg aacgcacacaag gtatagggcg 10740
gcgcctacaa tccatgccaatccatgttgcgtgcgg aggcggcata aatcgccgt 10800
acgatcagcg gtccaatgtat cgaagtttagg ctggtaagag cccgcagcga tccttgaagc 10860
tgtccctgtat ggtcgatc tacctgcctg gacagcatgg cctgcacacgc gggcatcccg 10920
atgcccgg aagcgagaag aatcataatg gggaaaggcca tccagcctcg cgtcgac 10980
gccagcaaga cgtagcccg cgcgtcgcc gccatgcgg cgataatggc ctgcgttc 11040
ccgaaacgtt tggggcgaccaggatgcg aaggcttgag cgagggcggtg caagattccg 11100
aataccgcaa gcgcacaggcc gatcatgcgc ggcgtccac gaaagcggtc ctcggcaaa 11160
atgaccggcaga gcgcgtgcgg cacctgttcc acgagttgca tgataaagaa gacagtcata 11220
agtgcggcga cgtatgtcat gccccgcgc caccggaaagg agctgactgg gttgaaggct 11280
ctcaaggggca tcggtcgagg atccttcaat atgcgcacat acgctgttat gttcaagggtc 11340

ctttcgttta agaacgaaaag cggcttcct tttgagggat gttcaagtt gttcaatct 11400
 atcaaatttg caaatccccca gtctgtatct agagcgttga atcggtgatg cgatttgttta 11460
 attaaattga tggtgtcacc attaccaggc cttagatatac caatggcaaa ctgagcacaa 11520
 caataccagt ccggatcaac tggcaccatc tctcccgtag tctcatctaa ttttcttcc 11580
 ggatgagggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11640
 atgtcccatt cgaaatcacc aattctaaac ctggcgaat tgtatttcgg gtttgttaac 11700
 tcggtccagt caggaatgtt ccacgtgaag ctatcttcca gcaaagtctc cacttcttca 11760
 tcaaattgtg gagaatactc ccaatgtct tatctatggg acttccgggaa aacacagtac 11820
 cgatacttcc caattcgtct tcagagctca ttgtttgtt gaagagacta atcaaagaat 11880
 cgtttctca aaaaaattaa tatcttaact gatagttga tcaaaggggc aaaacgtagg 11940
 ggcaaaacaaa cggaaaaatc gtttctcaaa ttctctgatg ccaagaactc taaccagtct 12000
 tatctaaaaa ttgccttatg atccgtctct ccgggtacag cctgtgttaac tgattaatcc 12060
 tgcctttcta atcaccattc taatgtttta attaaggat ttgtcttca ttaacggctt 12120
 tcgctcataa aaatgttatg acgtttgcc cgcaggcggg aaaccatcca cttcacgaga 12180
 ctgatctcct ctgccggaac accgggcatac tccaaacttat aagttggaga aataagagaa 12240
 ttccagattg agagaatgaa aaaaaaaaaac ccttagttca taggtccatt ctcttagcgc 12300
 aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtgtatg 12360
 caacctgcct ggagtaaatg atgacacaag gcaattgacc cacgcatgta tctatctcat 12420
 ttcttacac cttcttattac cttctgtct ctctgatttg gaaaaagctg aaaaaaaaaagg 12480
 ttgaaaccag ttccctgaaa ttattccct acttgactaa taagtatata aagacggtag 12540
 gtattgattg taattctgta aatcttatttc ttaaacttct taaattctac ttttatagtt 12600
 agtctttttt ttagtttaa aacaccaaga acttagttc gaataaacac acataaaca 12660
 acaagcttac aaaacaaa atg gct gca tat gca gct cag ggc tat aag gtg 12711
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val
 1 5 10
 cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac 12759
 Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr
 15 20 25
 atg tcc aag gct cat ggg atc gat cct aac atc agg acc ggg gtg aga 12807
 Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg
 30 35 40
 aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc 12855
 Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe
 45 50 55

ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys 60 65 70 75	12903
gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc att ggc act Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr 80 85 90	12951
gtc ctt gac caa gca gag act gcg ggg ggc aga ctg gtt gtg ctc gcc Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala 95 100 105	12999
acc gcc acc cct ccg ggc tcc gtc act gtg ccc cat ccc aac atc gag Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu 110 115 120	13047
gag gtt gct ctg tcc acc acc gga gag atc cct ttt tac ggc aag qct Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala 125 130 135	13095
atc ccc ctc gaa gta atc aag ggg ggg aga cat ctc atc ttc tgt cat Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His 140 145 150 155	13143
tca aag aag aag tgc gac gaa ctc gcc gca aag ctg gtc gca ttg ggc Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly 160 165 170	13191
atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg tcc gtc atc ccg Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro 175 180 185	13239
acc agc ggc gat gtt gtc gtc gtg gca acc gat gcc ctc atg acc ggc Thr Ser Gly Asp Val Val Val Ala Thr Asp Ala Leu Met Thr Gly 190 195 200	13287
tat acc ggc gac ttc gac tcg gtg ata gac tgc aat acg tgt gtc acc Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr 205 210 215	13335
cag aca gtc gat ttc agc ctt gac cct acc ttc acc att gag aca atc Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile 220 225 230 235	13383
acg ctc ccc caa gat gct gtc tcc cgc act caa cgt cgg ggc agg act Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr 240 245 250	13431
ggc agg ggg aag cca ggc atc tac aga ttt gtg gca ccg ggg gag cgc Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg 255 260 265	13479
ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gca Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala 270 275 280	13527
ggc tgt gct tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu 285 290 295	13575

cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat ctt 13623
 Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu
 300 305 310 315

gaa ttt tgg gag ggc gtc ttt aca ggc ctc act cat ata gat gcc cac 13671
 Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His
 320 325 330

ttt cta tcc cag aca aag cag agt ggg gag aac ctt cct tac ctg gta 13719
 Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val
 335 340 345

gcg tac caa gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg 13767
 Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser
 350 355 360

tgg gac cag atg tgg aag tgt ttg att cgc ctc aag ccc acc ctc cat 13815
 Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His
 365 370 375

ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa atc 13863
 Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile
 380 385 390 395

acc ctg acg cac cca gtc acc aaa tac atc atg aca tgc atg tcg gcc 13911
 Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala
 400 405 410

gac ctg gag gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg 13959
 Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu
 415 420 425

gct gct ttg gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata gtg 14007
 Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val
 430 435 440

ggc agg gtc gtc ttg tcc ggg aag ccg gca atc ata cct gac agg gaa 14055
 Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu
 445 450 455

gtc ctc tac cga gag ttc gat gag atg gaa gag tgc tct cag cac tta 14103
 Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu
 460 465 470 475

ccg tac atc gag caa ggg atg atg ctc gcc gag cag ttc aag cag aag 14151
 Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys
 480 485 490

gcc ctc ggc ctc ctg cag acc gcg tcc cgt cag gca gag gtt atc gcc 14199
 Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala
 495 500 505

cct gct gtc cag acc aac tgg caa aaa ctc gag acc ttc tgg gcg aag 14247
 Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys
 510 515 520

cat atg tgg aac ttc atc agt ggg ata caa tac ttg gcg ggc ttg tca 14295
 His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser
 525 530 535

acg ctg cct ggt aac ccc gcc att gct tca ttg atg gct ttt aca gct 14343
 Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala
 540 545 550 555

 gct gtc acc agc cca cta acc act agc caa acc ctc ctc ttc aac ata 14391
 Ala Val Thr Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile
 560 565 570

 ttg ggg ggg tgg gtg gct gcc cag ctc gcc gcc ccc ggt gcc gct act 14439
 Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr
 575 580 585

 gcc ttt gtg ggc gct ggc tta gct ggc gcc gcc atc ggc agt gtt gga 14487
 Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly
 590 595 600

 ctg ggg aag gtc ctc ata gac atc ctt gca ggg tat ggc gcg ggc gtg 14535
 Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val
 605 610 615

 gcg gga gct ctt gtg gca ttc aag atc atg agc ggt gag gtc ccc tcc 14583
 Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser
 620 625 630 635

 acg gag gac ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc gga gcc 14631
 Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala
 640 645 650

 ctc gta gtc ggc gtg gtc tgt gca gca ata ctg cgc cgg cac gtt ggc 14679
 Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly
 655 660 665

 ccg ggc gag ggg gca gtg cag tgg atg aac cgg ctg ata gcc ttc gcc 14727
 Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala
 670 675 680

 tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat 14775
 Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp
 685 690 695

 gca gct gcc cgc gtc act gcc ata ctc agc agc ctc act gta acc cag 14823
 Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln
 700 705 710 715

 ctc ctg agg cga ctg cac cag tgg ata agc tcg gag tgt acc act cca 14871
 Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro
 720 725 730

 tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg 14919
 Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val
 735 740 745

 ttg agc gac ttt aag acc tgg cta aaa gct aag ctc atg cca cag ctg 14967
 Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu
 750 755 760

 cct ggg atc ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg 15015
 Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp
 765 770 775

cga ggg gac ggc atc atg cac act cgc tgc cac tgt gga gct gag atc Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile 780 785 790 795	15063
act gga cat gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr 800 805 810	15111
tgc agg aac atg tgg agt ggg acc ttc ccc att aat gcc tac acc acg Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr 815 820 825	15159
ggc ccc tgt acc ccc ctt cct gcg ccg aac tac acg ttc gcg cta tgg Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp 830 835 840	15207
agg gtg tct gca gag gaa tac gtg gag ata agg cag gtg ggg gac ttc Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe 845 850 855	15255
cac tac gtg acg ggt atg act act gac aat ctt aaa tgc ccg tgc cag His Tyr Val Thr Gly Met Thr Asp Asn Leu Lys Cys Pro Cys Gln 860 865 870 875	15303
gtc cca tcg ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cat Val Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His 880 885 890	15351
agg ttt gcg ccc ccc tgc aag ccc ttg ctg cgg gag gag gta tca ttc Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe 895 900 905	15399
aga gta gga ctc cac gaa tac ccg gta ggg tcg caa tta cct tgc gag Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu 910 915 920	15447
ccc gaa ccg gac gtg gcc gtg ttg acg tcc atg ctc act gat ccc tcc Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser 925 930 935	15495
cat ata aca gca gag gcg gcc ggg cga agg ttg gcg agg gga tca ccc His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro 940 945 950 955	15543
ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc Pro Ser Val Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu 960 965 970	15591
aag gca act tgc acc gct aac cat gac tcc cct gat gct gag ctc ata Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile 975 980 985	15639
gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg 990 995 1000	15687
gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu 1005 1010 1015	15735

gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His 1260 1265 1270 1275	16503
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His 1280 1285 1290	16551
gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu 1295 1300 1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1310 1315 1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1325 1330 1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1340 1345 1350 1355	16743
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr 1360 1365 1370	16791
gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala 1375 1380 1385	16839
tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1390 1395 1400	16887
ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca atc Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile 1405 1410 1415	16935
tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser 1420 1425 1430 1435	16983
ctc acc gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly 1440 1445 1450	17031
gag aac tgc ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr 1455 1460 1465	17079
agc tgt ggt aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys 1470 1475 1480	17127
cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp 1485 1490 1495	17175

tta gtc gtt atc tgt gaa agc gcg ggg gtc cag gag gac gcg gcg agc Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser 1500 1505 1510 1515	17223
ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly 1520 1525 1530	17271
gac ccc cca caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser 1535 1540 1545	17319
tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr 1550 1555 1560	17367
ctc acc cgt gac cct aca acc ccc ctc gcg aga gct gcg tgg gag aca Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr 1565 1570 1575	17415
gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe 1580 1585 1590 1595	17463
gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser 1600 1605 1610	17511
gtc ctt ata gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile 1615 1620 1625	17559
tac ggg gcc tgc tac tcc ata gaa cca ctg gat cta cct cca atc att Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile 1630 1635 1640	17607
caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro 1645 1650 1655	17655
ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro 1660 1665 1670 1675	17703
ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu 1680 1685 1690	17751
ctg gcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac Leu Ala Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn 1695 1700 1705	17799
tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly 1710 1715 1720	17847
cag ctg gac ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac Gln Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp 1725 1730 1735	17895

att tat cac agc gtg tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc 17943
 Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys
 1740 1745 1750 1755

 cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga 17991
 Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1760 1765 1770

 atg agc acg aat cct aaa cct caa aga aag acc aaa cgt aac acc aac 18039
 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
 1775 1780 1785

 cg^g cg^g cc^g ca^g ga^c gtc aag tt^c cc^g gg^t gg^c gg^t ca^g atc gtt gg^t 18087
 Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
 1790 1795 1800

 gg^a gt^t ta^c tt^g tt^g cc^g cg^c ag^g gg^c cct ag^a tt^g gg^t gt^g cg^c gc^g 18135
 Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
 1805 1810 1815

 ac^g ag^a a^a g^c t^c g^a g^g cg^g tc^g ca^a c^c t^g ca^g a^a c^g t^c 18183
 Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro
 1820 1825 1830 1835

 at^c cc^c a^a g^c t^g cg^g cc^c g^a g^g gg^a cc^t t^g g^c t^g ca^g cc^c gg^t 18231
 Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly
 1840 1845 1850

 ta^c cc^t tt^g cc^c ct^c ta^t gg^c aa^t ga^g gg^c t^g cc^g gg^a tt^g 18279
 Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp
 1855 1860 1865

 ct^c ct^g t^c cc^c cg^t gg^c t^c cg^g cc^t ag^c t^g gg^c cc^c a^c g^c cc^c 18327
 Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
 1870 1875 1880

 cg^g cg^t ag^g tc^g cg^c a^a tt^g gg^t a^a g^g gtc at^c gat acc ct^t ac^g t^g 18375
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
 1885 1890 1895

 gg^c tt^c gg^c gac ct^c at^g gg^g ta^c at^a cc^g ct^c gtc taatagt^{cg}a 18421
 Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
 1900 1905 1910

 ct^ttg^tcc^c act^tgtact^t tag^tctcg^tac aaaatacaat atact^tttca tttctcc^{cg}ta 18481
 aacaacatgt tt^tcccatgt aatatcc^tttt tctat^tttt^t gttccgttac caact^ttaca 18541
 catactttat atagctattc acttctatac actaaaaaaac taagacaatt ttaat^tttgc 18601
 tgccctgccat atttcaattt gttataaattt cctataattt atcctatttag tagctaaaaaa 18661
 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721
 gccatgatcg cgtatcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781
 ccaaagcggt cggacagtgc tccgagaacg ggtgcgcata gaaattgcacat 18841
 agcgctagca gcacgcccata gtgactggcg atgctgtcgg aatggacgat atcccgcaag 18901

aggccggca gtacccggcat aaccaaggct atgcctacag catccagggt gacggtgccg 18961
 aggatgacga tgagcgcatt gtttagatttc atacacggtg cctgactgcg ttagcaattt 19021
 aactgtgata aactaccgca ttaaagctt ttctttccaa tttttttttt ttcgtcatta 19081
 taaaaatcat tacgaccgag attcccgggt aataactgat ataattaaat tgaagctcta 19141
 atttgtgagt ttagtataca tgcattttact tataatacag ttttttagtt ttgctggccg 19201
 catcttctca aatatgcttc ccagcctgct tttctgtaac gttcaccctc taccttagca 19261
 tcccttcctt ttgcaaatacg tcctttcca acaataataa tgtcagatcc tgtagagacc 19321
 acatcatcca cggttctata ctgttgaccc aatgcgtctc cttgtcatc taaacccaca 19381
 ccgggtgtca taatcaacca atcgtaacct tcatactcttc cacccatgtc tctttgagca 19441
 ataaagccga taacaaaatc tttgtcgctc ttgcataatgt caacagtacc ctttagtat 19501
 tctccagtag atagggagcc cttgcatacg aattctgcta acataaaaag gcctcttagt 19561
 tcctttgtta cttcttctgc cgccctgctt aaaccgctaa caatacctgg gcccaccaca 19621
 ccgtgtcat tcgtaatgtc tgccccattct gctattctgt atacacccgc agagtaactgc 19681
 aatttgactg tattaccaat gtcagcaat tttctgtt cgaagagtaa aaaattgtac 19741
 ttggcggata atgcctttag cggcttaact gtgccttcca tggaaaaatc agtcaagata 19801
 tccacatgtg ttttagtaa acaaattttg ggacctaattt cttcaactaa ctccagtaat 19861
 tccttggtgg tacgaacatc caatgaagca cacaagtttgc tttgttttc gtgcatagtata 19921
 ttaaatacgat tggcagcaac aggacttagga tgagtagcag cacgttcctt atatgtat 19981
 ttgcacatga tttatcttcg tttcctgcag gttttgttc tgtgcagtttgc gtttaagaat 20041
 actgggcaat ttcatgtttc ttcaacacta catatgcgtt tatataccaa tctaaatctg 20101
 tgctcccttcc ttctgttcc cttctgttcg gagattaccg aatcaaaaaa atttcaagga 20161
 aaccgaaatc aaaaaaaaaa aatgatgaat tgaaaagctt atcgat 20217

<210> 17
 <211> 1911
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core140

<400> 17
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
 1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
 20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
 35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
 50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser
 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
 85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
 100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
 115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
 130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys
 145 150 155 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala
 165 170 175

Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
 180 185 190

Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe
 195 200 205

Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe
 210 215 220

Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp
 225 230 235 240

Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro
 245 250 255

Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe
 260 265 270

Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr
 275 280 285

Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn
 290 295 300

Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly
 305 310 315 320

Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr
 325 330 335

Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr
 340 345 350

Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp
 355 360 365

Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu
 370 375 380

Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro
 385 390 395 400

Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val
 405 410 415

Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala
 420 425 430

Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu
 435 440 445

Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu
 450 455 460

Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln
 465 470 475 480

Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu
 485 490 495

Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
 500 505 510

Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe
 515 520 525

Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
 530 535 540

Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
 545 550 555 560

Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575

Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590

Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605

Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
 610 615 620

Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
 625 630 635 640

Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655

Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
 660 665 670
 Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
 675 680 685
 Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
 690 695 700
 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
 705 710 715 720
 His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735
 Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750
 Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765
 Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
 770 775 780
 Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
 785 790 795 800
 Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
 805 810 815
 Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
 820 825 830
 Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
 835 840 845
 Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
 850 855 860
 Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
 865 870 875 880
 Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
 885 890 895
 Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910
 Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
 915 920 925
 Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
 930 935 940
 Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
 945 950 955 960
 Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975

Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
 980 985 990

Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
 995 1000 1005

Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
 1010 1015 1020

Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
 025 1030 1035 1040

Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
 1045 1050 1055

Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070

Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
 1075 1080 1085

Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
 1090 1095 1100

Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
 105 1110 1115 1120

Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135

Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
 1140 1145 1150

Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
 1155 1160 1165

Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
 1170 1175 1180

Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 1185 1190 1195 1200

Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
 1205 1210 1215

Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
 1220 1225 1230

Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
 1235 1240 1245

Lys Glu Val Lys Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
 1250 1255 1260

Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 1265 1270 1275 1280

Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
 1285 1290 1295

Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310
 Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325
 Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 1330 1335 1340
 Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
 345 1350 1355 1360
 Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
 1365 1370 1375
 Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
 1380 1385 1390
 Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
 1395 1400 1405
 Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
 1410 1415 1420
 Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
 425 1430 1435 1440
 Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
 1445 1450 1455
 Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470
 Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
 1475 1480 1485
 Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
 1490 1495 1500
 Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
 505 1510 1515 1520
 Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535
 Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
 1540 1545 1550
 Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
 1555 1560 1565
 Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
 1570 1575 1580
 Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
 585 1590 1595 1600
 Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
 1605 1610 1615

Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
 1620 1625 1630

Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
 1635 1640 1645

Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
 1650 1655 1660

Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
 665 1670 1675 1680

Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
 1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser
 1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
 1730 1735 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala
 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro
 1765 1770 1775

Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
 1780 1785 1790

Val Lys Phe Pro Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu
 1795 1800 1805

Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
 1810 1815 1820

Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
 825 1830 1835 1840

Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
 1845 1850 1855

Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
 1860 1865 1870

Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
 1875 1880 1885

Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
 1890 1895 1900

Met Gly Tyr Ile Pro Leu Val
 905 1910

<211> 20247
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
pd.delta.NS3NS5.pj.core150

<220>
<221> CDS
<222> (12679)..(18441)

<400> 18
atcgatccta ccccttgcgc taaaagaagta tatgtgccta ctaacgcttg tctttgtctc 60
tgtcaactaaa cactggatta ttactcccag ataccttattt tggactaatt taaatgattt 120
cgatatcaacg ttcttaatat cgctgaatct tccacaattt atgaaaagtag ctaggaagag 180
gaattggtat aaagtttttg tttttgtaaa tctcgaagta tactcaaacg aatttagtat 240
tttctcagtg atctcccaga tgctttcacc ctcacttaga agtgctttaa gcattttttt 300
actgtggcta tttcccttat ctgcttcttc cgatgattcg aactgttaatt gcaaactact 360
tacaatatacgtt gtgatatcag attgatgttt ttgtccatag taaggaataa ttgtaaattc 420
ccaagcagga atcaatttct ttaatgaggc ttccagaatt gttgctttt gcgtcttgta 480
tttaaactgg agtattttat tgacaatatac gaaactcagc gaattgctta tgatagtatt 540
atagctcatg aatgtggctc tcttgattgc tttccgtta tgtgtaatca tccaacataa 600
ataggtagt tcagcagcac ataatgctat ttttcacccctt gaaggctttt caaaccttcc 660
cacaactga cgaacaagca ccttaggtgg ttttttacat aatataatcaa attgtggcat 720
gcttagcgcc gatcttgcgt gcaattgata tctagttca actactctat ttatcttgta 780
tcttgcatgtt ttcaaaacacg ctaactcgaa aaactaactt taattgtcctt gtttgcgtcg 840
cgttcttcg aaaaatgcac cggccgcgc ttatttgac tgcggaaaata attggactg 900
cggtatcttc atttcattt taaaaatgc acctttgctg cttttcctta attttttagac 960
ggccccgcagg ttctttgc ggtactatct tttttttttt acatgtgatc 1020
tgcacagatt ttataatgta ataagcaaga atacattatac aaacgaacaa tactggtaaa 1080
agaaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140
cttatacggtt ctggatgta tttttttttt ttattgaaat gattgctcctt gatgttagata 1200
ttgtatatacgtt gagtaaacgt aagtctgtt agtactctt tccaggatat gtcataaggc 1260
ccatggaaatc tctcacaacc ggtaggccgt atggcttga ttctagcgca gaagattcca 1320
gcgtatcttc tgactccagt gctgaggtaa ttttgctgc tgcgaagatg gtttaaggaaa 1380

ggtttatttc gattggaaat ggtatgctct cttcacaaga agcaagtcag gctgccatag 1440
atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500
ctataataat aggaagattg cccgagaaaag acaagaagag agctaccgaa atgctcatga 1560
aaaaaatgga ttgtacacag ttattagtcc caccagctcc aacggaagaa gatgttatga 1620
agctcgtaag cgtcgttacc caattgctta cttagtcc accagatcgt caagctgctt 1680
taataggta tttattcattc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740
tggcggcaga gaatcgttt cagcaaaaaa agagttagtt ggaaggaagg actgaagtga 1800
accatgctaa tacaaatgaa gaagttccct ccaggcgaac aagaagtaga gacacaatg 1860
caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacg 1920
aaaaaaggag agtagcaacg agggtaaggg gcagaaaaatc acgtaatact tctagggtat 1980
gatccaatat caaaggaaat gatagcattt aaggatgaga ctaatccat tgaggagtgg 2040
cagcatatag aacagctaaa gggtagtgct gaaggaagca tacgataccc cgcatggaat 2100
gggataatat cacaggaggt actagactac cttagtccct acataaaatag acgcatataa 2160
gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220
caacacgcag atatagggtc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280
tttcggaaag cgctcgaaaa cgaaaacgct ttgaagttcc tattccgaag ttcctattct 2340
ctagaaaatgta taggaacttc agagcgctt tgaaaaccaa aagcgctctg aagacgcact 2400
ttcaaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460
tccacaaaca ttgctaaaaa gatatcttt gctatatac tctgtgctat atccctatat 2520
aacctaccca tccacccctt gctccttcaa cttgcattca aactcgacct ctacatcaac 2580
aggcttccaa tgctcttcaa atttactgt caagtagacc catacggtcg taatatgctg 2640
ctcttcataa tgtaagctt tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700
ctttacgggtt ccctgagatt gaatttagttc cttagtata tgatacaaga cactttgaa 2760
ctttgtacga cgaattttga ggttcgcatt cctctggcta tttccaattt tcctgtcggc 2820
tattatctcc gcctcagttt gatctccgc ttcagactgc cattttcac ataatgaatc 2880
tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940
atgttgtaca ttgttttagtt cacgagaagg gtcctttca ggcggtagct cctgatctcc 3000
tatatgaccc ttatcctgtt ctctttccac aaacttagaa atgtattcat gaattatgga 3060
gcacctaata acattttca aggccggagaa gtttggccaa gatgcccaat atgcttgaca 3120
tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180

tcgaagataa gagaagaatg cagtgacatt tgtattgaca aatggagatt ccatgtatct 3240
aaaaaaatacg cctttaggcc ttctgatacc ctccccctg cggttagcg tgcctttac 3300
attaatatct aaaccctctc cgatggtgcc cttaactga ctaataaaatg caaccgatat 3360
aaactgtgat aattctgggt gatttatgat tcgatcgaca attgtattgt acactagtgc 3420
aggatcaggc caatccagtt cttttcaat taccgggtgtc tcgtctgtat tcagtacatg 3480
tccaacaaat gcaaattgcta acgttttgta tttcttataa ttgtcaggaa ctggaaaagt 3540
cccccttgtc gtctcgatta cacacactt ttcatcgta accataggtt ggaagtgctg 3600
cataatacat tgcttaatac aagcaagcag tctctcgcca ttcataatttc agttatttc 3660
cattacagct gatgtcattt gatatacgcg ctgtaaaaat ctatctgtta cagaaggttt 3720
tcgcgggttt tataaacaaa actttcgta cgaaatcgag caatcacccc agctgcgtat 3780
ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcata 3840
ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaac ctcaatctgt 3900
ccattgaatg ccttataaaaa cagctataga ttgcataagaa gagttagcta ctcaatgctt 3960
tttgtcaaaag cttactgatg atgatgtgtc tactttcagg cgggtctgtat gtaaggagaa 4020
tgacattata aagctggcac ttagaattcc acggactata gactatacta gtataactccg 4080
tctactgtac gatacacttc cgctcaggc tttgccttt aacgaggcct taccacttt 4140
ttgttactct attgatccag ctcagcaaag gcagtgtgat ctaagattct atcttcgcga 4200
tgttagaaaa ctagctagac cgagaaaagag actagaaaatg caaaaggcac ttctacaatg 4260
gctgccatca ttattatccg atgtgacgt gcattttttt tttttttttt tttttttttt 4320
ttttttttt tttttttttt ttttttgta caaatatcat aaaaaaaagag aatctttta 4380
agcaaggatt ttcttaactt cttcggcgac agcatcaccg acttcgggtgg tactgttgg 4440
accaccaaaa tcaccaggatc tgataacctgc atccaaaacc ttttaactg catctcaat 4500
ggcttacct tcttcaggca agttcaatga caattcaac atcattgcag cagacaagat 4560
agtggcgata gggttgacct tattcttgg caaatctgga gcggaaaccat ggcattgg 4620
gtacaaacca aatgcgggtgt tcttgcgtgg caaagaggcc aaggacgcag atggcaacaa 4680
acccaaggag cctgggataa cggaggcttc atcggagatg atatcaccaa acatgttgct 4740
ggtgattata ataccattta ggtgggttgg gttcttaactt aggtatcatgg cggcagaatc 4800
aatcaattga tggtaactt tcaatgttagg gaattcggtc ttgtatgggtt cctccacagt 4860
ttttctccat aatcttgaag aggccaaaac attagcttta tccaggacc aaataggcaa 4920
tggtggtca tggtaactt ccatgaaagc ggccattttt gtgattctt gcacttctgg 4980

aacggtgtat tgttcaactat cccaaagcgac accatcacca tcgtcttcct ttcttacc 5040
aaagtaaaata cctccccacta attctctaac aacaacgaag tcagtagcatt tagcaaattg 5100
tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5160
ggcgtacaat tgaagttctt tacggatttt tagtaaacct tgttcaggc taacactacc 5220
ggtaccccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttgaggc 5280
ttccagcgcc tcatctggaa gtggaacacc ttagcatcg atagcagcac caccaattaa 5340
atgattttcg aaatcgaact tgacattgga acgaacatca gaaatagctt taagaacctt 5400
aatggcttcg gctgtgattt cttgaccaac gtggtcaccc ggcaaaacga cgatcttctt 5460
aggggcagac attacaatgg tatatacctt aaatatataat aaaaaaaaaa aaaaaaaaaa 5520
aaaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacaggctaa 5580
tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640
acatccgaac ctgggagttt tccctgaaac agatagtata tttgaacctg tataataata 5700
tatagtctag cgctttacgg aagacaatgt atgtatttcg ttccctggag aaactattgc 5760
atctatttgca taggtaatct tgcacgtcgc atccccggtt cattttctgc gtttccatct 5820
tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttt tagaacaaaaa 5880
atgcaacgcg agagcgctaa ttttcaaac aaagaatctg agctgcattt ttacagaaca 5940
gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tttgtttcat ttttgaaaa 6000
caaaaatgca acgcgagagc gctaattttt caaacaaga atctgagctg catttttaca 6060
gaacagaaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttcttttttg 6120
ttctacaaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta gattactttt 6180
tttctccctt gtgcgtctta taatgcagtc tcttgataac tttttgcact gtaggtccgt 6240
taaggtaga agaaggctac tttgggtgtct attttctttt ccataaaaaaa agcctgactc 6300
caactcccgc gtttactgat tactagcaa gctgcgggtg catttttca agataaaggc 6360
atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6420
gcgtttaga ttcttcattt gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6480
atactacgta taggaaatgt ttacatttc gtattgtttt cgattcactc tatgaatagt 6540
tcttactaca attttttgc ctaaagagta atactagaga taaacataaa aaatgttagag 6600
gtcgagttt gatgcaagtt caaggagcga aagggtggatg ggttaggtat atagggatat 6660
agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag cggtattcgc 6720
aatattttag tagctcgta cagtccggtg cgttttgggt ttttgaaag tgcgtcttca 6780

gagcgctttt ggtttcaaa agcgctctga agttcctata ctttcttagag aataggaact 6840
tcggaatagg aacctcaaag cgttccgaa aacgagcgct tccgaaaatg caacgcgagc 6900
tgccacata cagctcactg ttcacgtcgc acctatatct gcgtgttgc tgtatatata 6960
tatacatgag aagaacggca tagtgcgtgt ttatgctta atgcgtactt atatgcgtct 7020
atttatgttag gatgaaaggt agtctagtac ctccctgtat attatccat tccatgcggg 7080
gtatcgtatg cttcccttcag cactaccctt tagctgttct atatgctgcc actcctcaat 7140
tggatttagtc tcatccttca atgctatcat ttcccttgat attggatcat atgcatacgta 7200
ccgagaaact agtgcgaagt agtgatcagg tattgctgtt atctgatgag tatacgttgt 7260
cctggccacg gcagaagcac gcttatcgct ccaatttccc acaacattag tcaactccgt 7320
taggccccttc attgaaagaa atgaggtcat caaatgtctt ccaatgtgag attttgggcc 7380
atttttata gcaaagattg aataaggcgc atttttcttc aaagctttat tgtacgatct 7440
gactaaggta tcttttaata attggatttc ctgttattt cttgaagaat tgccggctt 7500
atttactcgt ttttaggactg gttcagaatt cctcaaaaat tcataccaaat atacaagtgg 7560
atcgatgata agctgtcaaa catgagaatt cttgaagacg aaaggccctc gtgatacgcc 7620
tatTTTATA ggttaatgtc atgataataa tggTTTCTTA gacgtcagggt ggcactttc 7680
ggggaaatgt gcgcggAACc cctattgtt tattttcta aatacattca aatatgtatc 7740
cgctcatgag acaataaccc tgataaaatgc ttcaataata ttgaaaaagg aagagtatga 7800
gtattcaaca ttccgtgtc gcccTTattc cctttttgc ggcattttgc ttccctgttt 7860
ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7920
tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagttt cgcggcaag 7980
aacgtttcc aatgatgagc acttttaaag ttctgctatg tggcgccgtt ttatccctgt 8040
ttgacgcccgg gcaagagcaa ctcggTCGCC gcatacacta ttctcagaat gacttggttg 8100
agtactcacc agtcacagaa aagcatctt cggatggcat gacagtaaga gaattatgca 8160
gtgctgccccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcgag 8220
gaccgaagga gctaaccgct ttttgcaca acatggggga tcatgtaact cgccttgatc 8280
gttgggaacc ggagctgaat gaagccatac caaAcgcga gCGTgacacc acgatgcctg 8340
cagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagttccc 8400
ggcaacaatt aatagactgg atggaggcgg ataaaagttgc aggaccactt ctgcgctcgg 8460
cccttccggc tggctggttt attgctgata aatctggage cggtgagcgt gggctcgcg 8520
gtatcattgc agcaactgggg ccagatggta agccctcccg tATCgttagtt atctacacqa 8580

cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640
tgattaagca ttgtaactg tcagaccaag tttactcata tatacttttag attgatttaa 8700
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgcataat ctcatgacca 8760
aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820
gatcttccttgc agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaaccac 8880
cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactctttt ccgaaggtaa 8940
ctggcttcag cagagcgcag ataccaaata ctgtccttct agttagccg tagttaggcc 9000
accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaattc ctgttaccag 9060
tggctgctgc cagtgccgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9120
cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180
gaacgaccta caccgaactg agataacctac agcgtgagct atgagaaaagc gccacgcttc 9240
ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 9300
cgagggagct tccagggggaa aacgccttgtt atctttatag tccctgtcggg ttccgcacc 9360
tctgacttga gcgtcgattt ttgtgatgtc cgtcaggggg gcggagccta tgaaaaacg 9420
ccagcaacgc ggcctttta cggttcctgg cctttgctg gcctttgct cacatgttct 9480
ttcctgcgtt atcccctgat tctgtggata accgttattac cgccttgag tgagctgata 9540
ccgctcgccg cagccgaacg accgagcgcgca gcgagtcagt gagcgaggaa gcggaaagagc 9600
gcctgatgcg gtatttctc cttacgcattc tgtgcggtat ttcacaccgc atatggtgca 9660
ctctcagttac aatctgtctt gatgccgcattt agttaagccca gtatacactc cgctatcgct 9720
acgtgactgg gtcattggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780
ggcttgcgtc ctcccgcat cgcattacag acaagctgtc accgtctccg ggagctgcat 9840
gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctgcgtt aaagctcatc 9900
agcgtggctg tgaagcgatt cacagatgtc tgcctgtca tccgcgtcca gtcgtttag 9960
tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020
ttcctgtttt gtcactgtatg cctccgtgtt aaaaaaaaaaaaaatgatgatgat 10080
accgatgaaa cgagagagga tgctcacat acgggttact gatgtgaac atgcccgggtt 10140
actggAACGT tttttttttt aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200
caactcagggtt caatgccagc gcttcgttaa tacagatgtt ggtgtccac aggtagccaa 10260
gcagcatcct gcgtgcaga tccggaaacat aatggtgcag ggctgtact tccgcgtttc 10320
cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380

tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10440
aaggcaaccc cgccagccta gccgggtcct caacgacagg agcacgatca tgcgcacccg 10500
tggccaggac ccaacgctgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10560
gatggatatg ttctgccaag ggttggtttgcgcattcaca gttctccgca agaattgatt 10620
ggctccaatt ctggagtgaa tgaatccgtt agcgaggtgc cgccggcttc cattcaggc 10680
gaggtggccc ggctccatgc accgcgacgc aacgcggggaa ggcagacaag gtatagggcg 10740
gcccctacaa tccatgccaa cccgttccat gtgctcgccg aggccgcata aatcgccgt 10800
acgatcagcg gtccaatgat cgaagtttagg ctggtaagag ccgcgagcga tccttgaagc 10860
tgtccctgat ggtcgcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10920
atgcccggg aagcgagaag aatcataatg gggaaaggcca tccagcctcg cgtcgcgaac 10980
gccagcaaga cgtagccccag cgctcgccgccc gccatgcccgg cgataatggc ctgcttctcg 11040
ccgaaacgtt tggggcggg accagtgacg aaggcttgag cgagggcgtg caagattccg 11100
aataccgcaa gcgacaggcc gatcatcgctc ggcgtccagc gaaagcggc ctcgccgaaa 11160
atgaccaga gcgctgcccgg cacctgtcct acgagttgca tgataaagaa gacagtcata 11220
agtgcggcga cgatagtcat gccccgcgcc caccggaaagg agctgactgg gttgaaggct 11280
ctcaaggcga tcggtcgagg atccctcaat atgcgcacat acgctgttat gttcaaggc 11340
ccttcgttta agaacgaaag cggcttcct tttgagggat gttcaagtt gttcaaatct 11400
atcaaatttg caaatccccca gtctgtatct agagcgttga atcggtgatg cgatttgtta 11460
attaaattga tggtgtcacc attaccaggt cttagatatac caatggcaaa ctgagcaca 11520
caataccagt ccggatcaac tggcaccatc tctccgtag tctcatctaa ttttcttcc 11580
ggatgaggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11640
atgtcccatt cgaaatcacc aattctaaac ctgggcgaat tgtatccgg gtttgttaac 11700
tcgttccagt caggaatgtt ccacgtgaag ctatcttcca gcaaagtctc cacttcttca 11760
tcaaattgtg gagaatactc ccaatgtct tatctatggg acttccggaa aacacagtac 11820
cgatacttcc caattcgctc tcagagctca ttgtttgtttt gaagagacta atcaaagaat 11880
cgttttctca aaaaaattaa tatcttaact gatagtttga tcaaagggc aaaacgtagg 11940
ggcaaacaaa cggaaaaatc gtttctcaaa ttttctgatg ccaagaactc taaccagtct 12000
tatctaaaaa ttgccttatg atccgtctct ccggttacag cctgtgttac 12060
tgcctttctca atcaccatc taatgtttta attaagggat tttgtttca ttaacggctt 12120
tcgctcataa aaatgttatg acgtttgcc cgcaggcggg aaaccatcca ctgcacgaga 12180

ctgatctcct ctgccggAAC accgggcATC tccaaCTTAT aagttggAGA aataagAGAA 12240
 ttTCAGATTG agagaATGAA aaaaaaaaaAC cCTTAGTTCA taggtCCATT ctCTTAGCGC 12300
 aactacAGAG AACAGGGGCA caaacAGGCa AAAAACGGGC acaacCTCAA tggagtGTAT 12360
 caacCTGCCT ggAGTAAATG atgacacaAG gcaattGACC cacGcatGTA tCTATCTCAT 12420
 tttCTTACAC CTTCTATTAC CTTCTGTCT CTCTGATTTG gaaaaAGCTG aaaaaaaaAGG 12480
 ttGAAACCAG ttCCCTGAAA ttattCCCT acTTGACTAA taAGTATA aAGACGGTAG 12540
 gtattGATTG taattCTGTA aatCTATTc ttaAAACTCT taaATTCTAC ttttATAGTT 12600
 agtCTTTTT ttagTTTAA aacaccaAGA acTTAGTTc gaataAAACAC acataAAACAA 12660
 acaagCTTAC AAAACAAATG GCT GCA TAT GCA GCT CAG GGC TAT AAG GTG 12711
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val
 1 5 10

cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac 12759
 Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr
 15 20 25

atg tcc aag gct cat ggg atc gat CCT AAC ATC AGG ACC GGG GTG AGA 12807
 Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg
 30 35 40

aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc 12855
 Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe
 45 50 55

ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt 12903
 Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys
 60 65 70 75

gac gag tgc cac tcc acg gat GCC ACA TCC ATC TTG GGC ATT GGC ACT 12951
 Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr
 80 85 90

gtc ctt gaccaa gca gag act GCG GGG GCG AGA CTG GTT GTG CTC GCC 12999
 Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala
 95 100 105

acc GCC ACC CCT CCg GGC TCC GTC ACT GTG CCC CAT CCC AAC ATC GAG 13047
 Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu
 110 115 120

gag gtt gct ctg tcc acc acc gga gag atc CCT TTT Tac GGC AAG GCT 13095
 Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala
 125 130 135

atc CCC CTC GAA Gta ATC AAG GGG GGG AGA CAT CTC ATC TTC TGT CAT 13143
 Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His
 140 145 150 155

tca aag aag aag tgc gac gaa CTC GCC GCA AAG CTG GTC GCA TTG GGC 13191
 Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly
 160 165 170

atc aat gcc gtc gcc tac tac cgc ggt ctt gac gtc gtg tcc gtc atc ccg Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro 175 180 185	13239
acc agc ggc gat gtt gtc gtc gtg gca acc gat gcc ctc atg acc ggc Thr Ser Gly Asp Val Val Val Ala Thr Asp Ala Leu Met Thr Gly 190 195 200	13287
tat acc ggc gac ttc gac tcg gtg ata gac tgc aat acg tgt gtc acc Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr 205 210 215	13335
cag aca gtc gat ttc agc ctt gac cct acc ttc acc att gag aca atc Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile 220 225 230 235	13383
acg ctc ccc caa gat gct gtc tcc cgc actcaa cgt cgg ggc agg act Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr 240 245 250	13431
ggc agg ggg aag cca ggc atc tac aga ttt gtg gca ccg ggg gag cgc Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg 255 260 265	13479
ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gca Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala 270 275 280	13527
ggc tgt gct tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu 285 290 295	13575
cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat ctt Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu 300 305 310 315	13623
gaa ttt tgg gag ggc gtc ttt aca ggc ctc act cat ata gat gcc cac Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His 320 325 330	13671
ttt cta tcc cag aca aag cag agt ggg gag aac ctt cct tac ctg gta Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val 335 340 345	13719
gcg tac caa gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser 350 355 360	13767
tgg gac cag atg tgg aag tgt ttg att cgc ctc aag ccc acc ctc cat Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His 365 370 375	13815
ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa atc Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile 380 385 390 395	13863
acc ctg acg cac cca gtc acc aaa tac atc atg aca tgc atg tcg gcc Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala 400 405 410	13911

gac ctg gag gtc gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu 415 420 425	13959
gct gct ttg gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata gtc Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val 430 435 440	14007
ggc agg gtc gtc ttg tcc ggg aag ccg gca atc ata cct gac agg gaa Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu 445 450 455	14055
gtc ctc tac cga gag ttc gat gag atg gaa gag tgc tct cag cac tta Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu 460 465 470 475	14103
ccg tac atc gag caa ggg atg atg ctc gcc gag cag ttc aag cag aag Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys 480 485 490	14151
gcc ctc ggc ctc ctg cag acc gcg tcc cgt cag gca gag gtt atc gcc Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala 495 500 505	14199
cct gct gtc cag acc aac tgg caa aaa ctc gag acc ttc tgg gcg aag Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys 510 515 520	14247
cat atg tgg aac ttc atc agt ggg ata caa tac ttg gcg ggc ttg tca His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser 525 530 535	14295
acg ctg cct ggt aac ccc gcc att gct tca ttg atg gct ttt aca gct Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala 540 545 550 555	14343
gct gtc acc agc cca cta acc act agc caa acc ctc ctc ttc aac ata Ala Val Thr Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile 560 565 570	14391
ttg ggg ggg tgg gtg gct gcc cag ctc gcc gcc ccc ggt gcc gct act Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr 575 580 585	14439
gcc ttt gtg ggc gct ggc tta gct ggc gcc atc ggc agt gtt gga Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly 590 595 600	14487
ctg ggg aag gtc ctc ata gac atc ctt gca ggg tat ggc gcg ggc gtg Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val 605 610 615	14535
gcg gga gct ctt gtg gca ttc aag atc atg agc ggt gag gtc ccc tcc Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser 620 625 630 635	14583
acg gag gac ctg gtc aat cta ctg ccc gcc atc ctc tcg ccc gga gcc Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala 640 645 650	14631

ctc gta gtc ggc gtg gtc tgt gca gca ata ctg cgc cgg cac gtt ggc Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly 655 660 665	14679
ccg ggc gag ggg gca gtg cag tgg atg aac cgg ctg ata gcc ttc gcc Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala 670 675 680	14727
tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp 685 690 695	14775
gca gct gcc cgc gtc act gcc ata ctc agc agc ctc act gta acc cag Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln 700 705 710 715	14823
ctc ctg agg cga ctg cac cag tgg ata agc tcg gag tgt acc act cca Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro 720 725 730	14871
tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val 735 740 745	14919
ttg agc gac ttt aag acc tgg cta aaa gct aag ctc atg cca cag ctg Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu 750 755 760	14967
cct ggg atc ccc ttt gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp 765 770 775	15015
cga ggg gac ggc atc atg cac act cgc tgc cac tgt gga gct gag atc Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile 780 785 790 795	15063
act gga cat gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr 800 805 810	15111
tgc agg aac atg tgg agt ggg acc ttc ccc att aat gcc tac acc acg Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr 815 820 825	15159
ggc ccc tgt acc ccc ctt cct gcg ccg aac tac acg ttc gcg cta tgg Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp 830 835 840	15207
agg gtg tct gca gag gaa tac gtg gag ata agg cag gtg ggg gac ttc Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe 845 850 855	15255
cac tac gtg acg ggt atg act act gac aat ctt aaa tgc ccg tgc cag His Tyr Val Thr Gly Met Thr Asp Asn Leu Lys Cys Pro Cys Gln 860 865 870 875	15303
gtc cca tcg ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cat Val Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His 880 885 890	15351

agg ttt gcg ccc ccc tgc aag ccc ttg ctg cg_g gag g_g g_a gta tca ttc 15399
 Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe
 895 900 905

aga gta gga ctc cac gaa tac ccg gta ggg tcg caa tta cct tgc gag 15447
 Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu
 910 915 920

ccc gaa ccg gac gtg gcc gtg ttg acg tcc atg ctc act gat ccc tcc 15495
 Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser
 925 930 935

cat ata aca gca gag g_c g_g g_g cga agg ttg g_c g_g agg gga tca ccc 15543
 His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro
 940 945 950 955

ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc 15591
 Pro Ser Val Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu
 960 965 970

aag gca act tgc acc gct aac cat gac tcc cct gat gct gag ctc ata 15639
 Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile
 975 980 985

gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg 15687
 Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg
 990 995 1000

gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt 15735
 Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu
 1005 1010 1015

gtg g_c g_g g_g g_a g_g c_g g_g atc tcc gta ccc gca gaa atc ctg 15783
 Val Ala Glu Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu
 1020 1025 1030 1035

cg_g aag tct cgg aga ttc gcc cag gcc ctg ccc gtt tgg g_c g_g c_g c_g
 Arg Lys Ser Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro 15831
 1040 1045 1050

gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa 15879
 Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu
 1055 1060 1065

cca cct gtg gtc cat ggc tgc cc_g ctt cca cct cca aag tcc cct cct 15927
 Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro
 1070 1075 1080

gtg cct ccg cct cgg aag aag cgg acg gtg gtc ctc act gaa tca acc 15975
 Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr
 1085 1090 1095

cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc
 Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser 16023
 1100 1105 1110 1115

tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc 16071
 Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro
 1120 1125 1130

gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser	1135	1140	1145	16119
atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly	1150	1155	1160	16167
tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys	1165	1170	1175	16215
tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala	1180	1185	1190	1195
gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu	1200	1205	1210	16311
cgt cac cac aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln	1215	1220	1225	16359
agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His	1230	1235	1240	16407
tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ser Lys Val Lys	1245	1250	1255	16455
gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His	1260	1265	1270	1275
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His	1280	1285	1290	16551
gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu	1295	1300	1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu	1310	1315	1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu	1325	1330	1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu	1340	1345	1350	1355
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr	1360	1365	1370	16791

gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala 1375 1380 1385	16839
tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1390 1395 1400	16887
ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca atc Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Ala Ile 1405 1410 1415	16935
tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser 1420 1425 1430 1435	16983
ctc acc gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly 1440 1445 1450	17031
gag aac tgc ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr 1455 1460 1465	17079
agc tgt ggt aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys 1470 1475 1480	17127
cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp 1485 1490 1495	17175
tta gtc gtt atc tgt gaa agc gcg ggg gtc cag gag gac gcg gcg agc Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser 1500 1505 1510 1515	17223
ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly 1520 1525 1530	17271
gac ccc cca caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser 1535 1540 1545	17319
tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr 1550 1555 1560	17367
ctc acc cgt gac cct aca acc ccc ctc gcg aga gct gcg tgg gag aca Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr 1565 1570 1575	17415
gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe 1580 1585 1590 1595	17463
gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser 1600 1605 1610	17511

gtc ctt ata gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile 1615 1620 1625	17559
tac ggg gcc tgc tac tcc ata gaa cca ctg gat cta cct cca atc att Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile 1630 1635 1640	17607
caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro 1645 1650 1655	17655
ggg gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro 1660 1665 1670 1675	17703
ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu 1680 1685 1690	17751
ctg gcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac Leu Ala Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn 1695 1700 1705	17799
tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly 1710 1715 1720	17847
cag ctg gac ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac Gln Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp 1725 1730 1735	17895
att tat cac agc gtg tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys 1740 1745 1750 1755	17943
cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg 1760 1765 1770	17991
atg agc acg aat cct aaa cct caa aga aag acc aaa cgt aac acc aac Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn 1775 1780 1785	18039
cgg cgg ccg cag gac gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gln Ile Val Gly 1790 1795 1800	18087
gga gtt tac ttg ttg ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala 1805 1810 1815	18135
acg aga aag act tcc gag cgg tcg caa cct cga ggt aga cgt cag cct Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro 1820 1825 1830 1835	18183
atc ccc aag gct cgt cgg ccc gag ggc agg acc tgg gct cag ccc ggg Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly 1840 1845 1850	18231

tac cct tgg ccc ctc tat ggc aat gag ggc tgc ggg tgg gcg gga tgg 18279
 Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp
 1855 1860 1865

ctc ctg tct ccc cgt ggc tct cgg cct agc tgg ggc ccc aca gac ccc 18327
 Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
 1870 1875 1880

cg^g cgt agg tcg cgc aat ttg ggt aag gtc atc gat acc ctt acg tgc 18375
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
 1885 1890 1895

ggc ttc gcc gac ctc atg ggg tac ata ccg ctc gtc ggc gcc cct ctt 18423
 Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu
 1900 1905 1910 1915

gga ggc gct gcc agg gcc taatagtcga ctttgttccc actgtacttt 18471
 Gly Gly Ala Ala Arg Ala
 1920

tagctcgta aaaaatacaat atactttca tttctccgt aacaacatgt tttccatgt 18531
 aatatccccc tctatcccc gttccgttac caactttaca catactttat atagctattc 18591
 acttctatac actaaaaaaac taagacaatt ttaatccc tgccctgccat atttcaattt 18651
 gttataaaatt cctataattt atcctatttag tagctaaaaa aagatgaatg tgaatcgaat 18711
 cctaagagaa ttggatctga tccacaggac ggggtggc gccatgatcg cgtagtcgat 18771
 agtggctcca agtagcgaag cgagcaggac tggcgccgg ccaaagcggt cggacagtgc 18831
 tccgagaacg ggtgcgcata gaaattgcata caacgcata agcgctagca gcacgccata 18891
 gtgactggcg atgctgtcg aatggacgat atcccgcaag aggccggca gtaccggcat 18951
 aaccaagcct atgcctacag catccagggt gacggtgccg aggatgacga tgagcgcatt 19011
 gttagatttc atacacggtg cctgactgcg ttagcaattt aactgtgata aactaccgca 19071
 ttaaagcttt ttcttccaa ttttttttt ttcgtcatta taaaaatcat tacgaccgag 19131
 atccccgggt aataactgat ataattaaat tgaagctcta atttgcgtt ttagtataca 19191
 tgcatttact tataatacag ttttttagtt ttgctggccg catcttctca aatatgcctc 19251
 ccagcctgct ttctgttaac gttcacccctc taccttagca tccctccct ttgcaaata 19311
 tcctcttcca acaataataa tgtcagatcc tgtagagacc acatcatcca cggttctata 19371
 ctgttgaccc aatgcgtctc ccttgcatac taaacccaca cgggtgtca taatcaacca 19431
 atcgtaacct tcatctcttc cacccatgtc tcttgagca ataaagccga taacaaaatc 19491
 tttgtcgctc ttgcataatgt caacagtacc ctttagtatat tctccagtag atagggagcc 19551
 cttgcatac aattctgcta acatcaaaag gcctcttaggt tcctttgtta cttcttctgc 19611
 cgcctgcttc aaaccgctaa caatacctgg gcccaccaca cgggtgtcat tcgtaatgtc 19671

tgcccattct gctattctgt atacaccgcg agagtaactgc aatttgactg tattaccaat 19731
 gtcagcaaat ttctgtctt cgaagagata aaaattgtac ttggcggata atgccttag 19791
 cggcttaact gtgccctcca tggaaaaatc agtcaagata tccacatgtg ttttagtaa 19851
 acaaatttg ggacctaattc cttcaactaa ctccagtaat tccttggtgg tacgaacatc 19911
 caatgaagca cacaagtttgc ttgcattttc gtgcattgata taaaatagct tggcagcaac 19971
 aggacttagga ttagtgcag cacgttcctt atatgttagct ttgcacatgat 20031
 ttccctgcag gttttgttc tgcattttc ggttaagaat actgggcaat ttcatgtttc 20091
 ttcaacacta catatgcgta tatataccaa tctaagtctg tgctccttcc ttcatgtttc 20151
 cttctgttcg gagattaccg aatcaaaaaa atttcaagga aaccgaaatc aaaaaaaaaaaga 20211
 ataaaaaaaaa aatgatgaat tgaaaagctt atcgat 20247

<210> 19
 <211> 1921
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core150

<400> 19
 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
 1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
 20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
 35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
 50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Cys Asp Glu Cys His Ser
 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
 85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
 100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
 115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
 130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys

145	150	155	160
Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala			
165	170	175	
Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val			
180	185	190	
Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe			
195	200	205	
Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe			
210	215	220	
Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp			
225	230	235	240
Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro			
245	250	255	
Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe			
260	265	270	
Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr			
275	280	285	
Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn			
290	295	300	
Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly			
305	310	315	320
Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr			
325	330	335	
Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr			
340	345	350	
Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp			
355	360	365	
Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu			
370	375	380	
Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro			
385	390	395	400
Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val			
405	410	415	
Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala			
420	425	430	
Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu			
435	440	445	
Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu			
450	455	460	

Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln
 465 470 475 480

Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu
 485 490 495

Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
 500 505 510

Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe
 515 520 525

Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
 530 535 540

Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
 545 550 555 560

Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575

Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590

Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605

Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
 610 615 620

Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
 625 630 635 640

Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655

Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
 660 665 670

Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
 675 680 685

Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
 690 695 700

Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
 705 710 715 720

His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735

Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750

Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765

Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
 770 775 780

Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
 785 790 795 800

 Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
 805 810 815

 Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
 820 825 830

 Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
 835 840 845

 Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
 850 855 860

 Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
 865 870 875 880

 Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
 885 890 895

 Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910

 Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
 915 920 925

 Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
 930 935 940

 Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
 945 950 955 960

 Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975

 Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
 980 985 990

 Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
 995 1000 1005

 Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
 1010 1015 1020

 Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
 1025 1030 1035 1040

 Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
 1045 1050 1055

 Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070

 Gly Cys Pro Leu Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
 1075 1080 1085

 Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
 1090 1095 1100

Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
 105 1110 1115 1120
 Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135
 Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
 1140 1145 1150
 Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
 1155 1160 1165
 Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
 1170 1175 1180
 Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 1185 1190 1195 1200
 Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
 1205 1210 1215
 Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
 1220 1225 1230
 Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
 1235 1240 1245
 Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
 1250 1255 1260
 Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 1265 1270 1275 1280
 Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
 1285 1290 1295
 Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310
 Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325
 Pro Glu Lys Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 1330 1335 1340
 Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
 1345 1350 1355 1360
 Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
 1365 1370 1375
 Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
 1380 1385 1390
 Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
 1395 1400 1405
 Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
 1410 1415 1420

Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
425 1430 1435 1440

Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
1445 1450 1455

Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
1460 1465 1470

Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
1475 1480 1485

Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
1490 1495 1500

Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
505 1510 1515 1520

Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
1525 1530 1535

Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
1540 1545 1550

Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
1555 1560 1565

Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
1570 1575 1580

Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
585 1590 1595 1600

Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
1605 1610 1615

Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
1620 1625 1630

Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
1635 1640 1645

Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
1650 1655 1660

Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
665 1670 1675 1680

Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser
1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
1730 1735 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala
745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro
1765 1770 1775

Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
1780 1785 1790

Val Lys Phe Pro Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu
1795 1800 1805

Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
1810 1815 1820

Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
825 1830 1835 1840

Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
1845 1850 1855

Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
1860 1865 1870

Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
1875 1880 1885

Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
1890 1895 1900

Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Ala Ala Arg
905 1910 1915 1920

Ala

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.